

# 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")[1934:2047]
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	(
1934	1934	EAST RAJASTHAN	1903	1.9	0.7	1.3	0.1	12.9	15.6	238.2	229.1	168.5	
1935	1935	EAST RAJASTHAN	1904	4.3	5.5	21.7	0.2	27.5	49.9	289.7	223.5	50.2	
1936	1936	EAST RAJASTHAN	1905	4.1	8.8	3.2	1.6	2.0	14.4	130.5	30.9	83.8	
1937	1937	EAST RAJASTHAN	1906	0.2	23.1	6.2	0.0	1.6	61.9	262.5	95.5	191.4	
1938	1938	EAST RAJASTHAN	1907	6.3	35.7	7.0	14.3	12.7	18.5	134.1	319.8	3.3	
...	...	...	...	...	...	...	...	...	...	...	...	...	
2042	2042	EAST RAJASTHAN	2011	0.0	11.2	0.2	0.5	5.1	140.9	193.6	284.1	166.4	
2043	2043	EAST RAJASTHAN	2012	1.9	0.0	0.0	3.6	9.5	11.2	170.5	365.0	131.3	
2044	2044	EAST RAJASTHAN	2013	1.4	21.7	0.4	3.2	1.0	90.6	319.0	278.5	88.0	
2045	2045	EAST RAJASTHAN	2014	28.4	10.0	6.4	7.3	8.4	23.5	197.1	261.0	136.9	
2046	2046	EAST RAJASTHAN	2015	12.1	0.1	55.9	15.9	3.5	96.4	297.6	142.8	20.1	

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	OCT
1934	1934	EAST RAJASTHAN	1903	1.9	0.7	1.3	0.1	12.9	15.6	238.2	229.1	168.5	10.2
1935	1935	EAST RAJASTHAN	1904	4.3	5.5	21.7	0.2	27.5	49.9	289.7	223.5	50.2	1.3
1936	1936	EAST RAJASTHAN	1905	4.1	8.8	3.2	1.6	2.0	14.4	130.5	30.9	83.8	1.3
1937	1937	EAST RAJASTHAN	1906	0.2	23.1	6.2	0.0	1.6	61.9	262.5	95.5	191.4	1.3
1938	1938	EAST RAJASTHAN	1907	6.3	35.7	7.0	14.3	12.7	18.5	134.1	319.8	3.3	1.3
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2042	2042	EAST RAJASTHAN	2011	0.0	11.2	0.2	0.5	5.1	140.9	193.6	284.1	166.4	1.3
2043	2043	EAST RAJASTHAN	2012	1.9	0.0	0.0	3.6	9.5	11.2	170.5	365.0	131.3	1.3
2044	2044	EAST RAJASTHAN	2013	1.4	21.7	0.4	3.2	1.0	90.6	319.0	278.5	88.0	1.3
2045	2045	EAST RAJASTHAN	2014	28.4	10.0	6.4	7.3	8.4	23.5	197.1	261.0	136.9	1.3
2046	2046	EAST RAJASTHAN	2015	12.1	0.1	55.9	15.9	3.5	96.4	297.6	142.8	20.1	1.3

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, 1934 to 2046
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
```

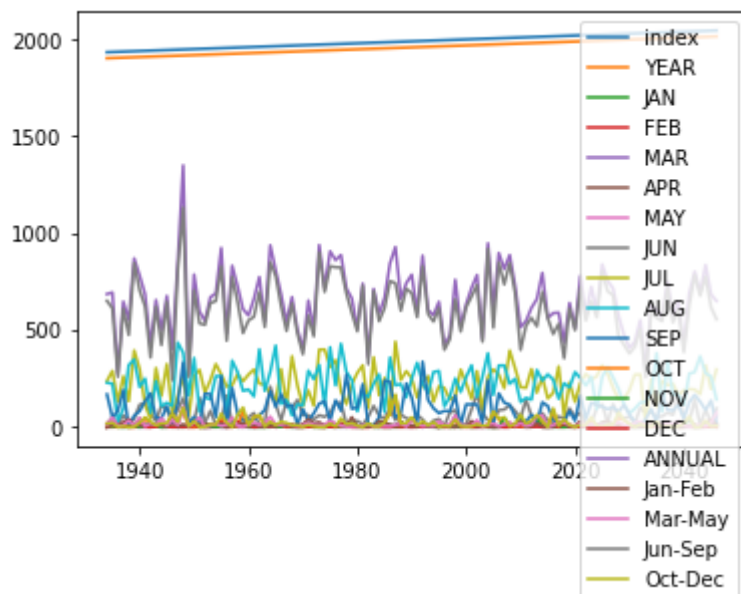
# Line Chart

In [6]:

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

Out[6]:

&lt;AxesSubplot:&gt;



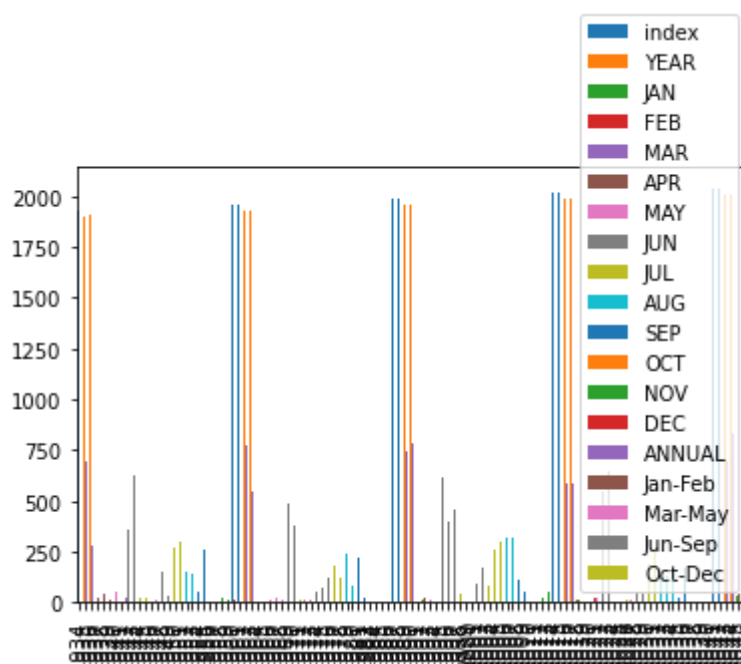
## Bar chart

In [7]:

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

Out[7]:

&lt;AxesSubplot:&gt;



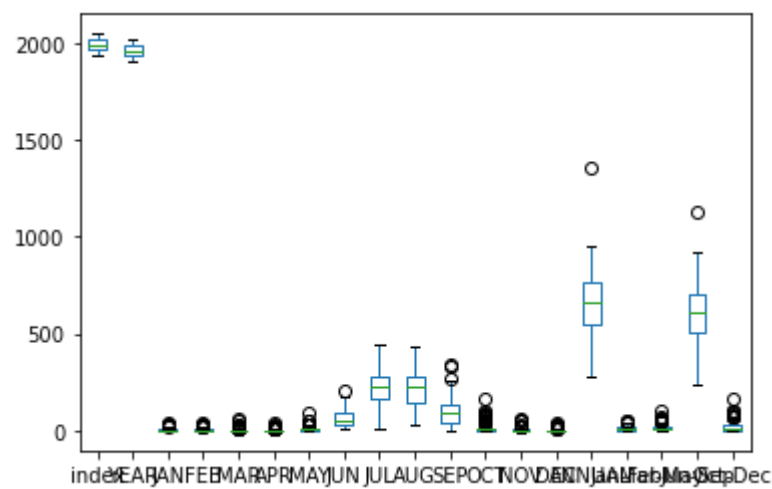
# Box chart

In [8]:

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

Out[8]:

<AxesSubplot:>



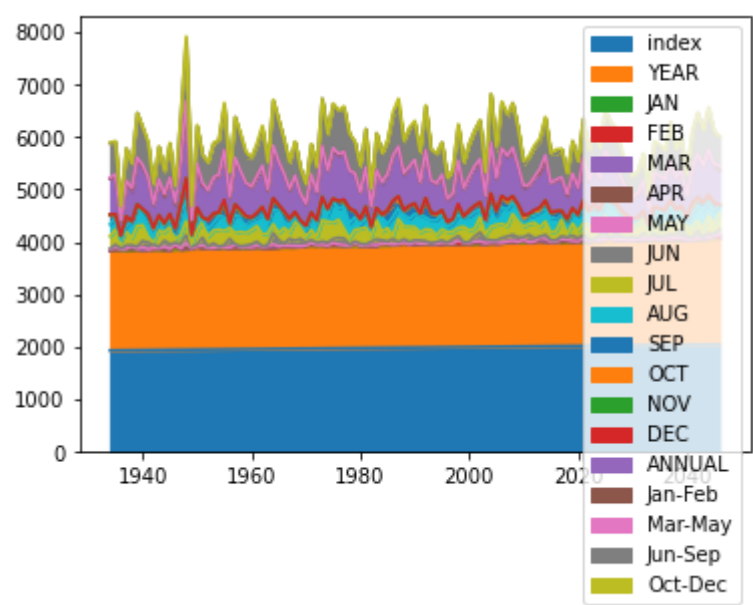
# Area Chart

In [9]:

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

Out[9]:

<AxesSubplot:>



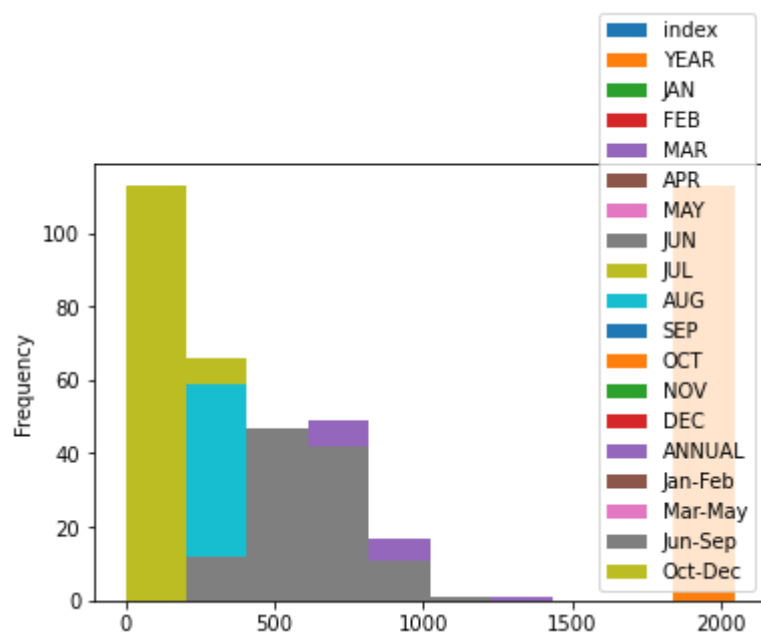
# Histogram

In [10]:

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

Out[10]:

<AxesSubplot:ylabel='Frequency'>



pie chart

In [11]:

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

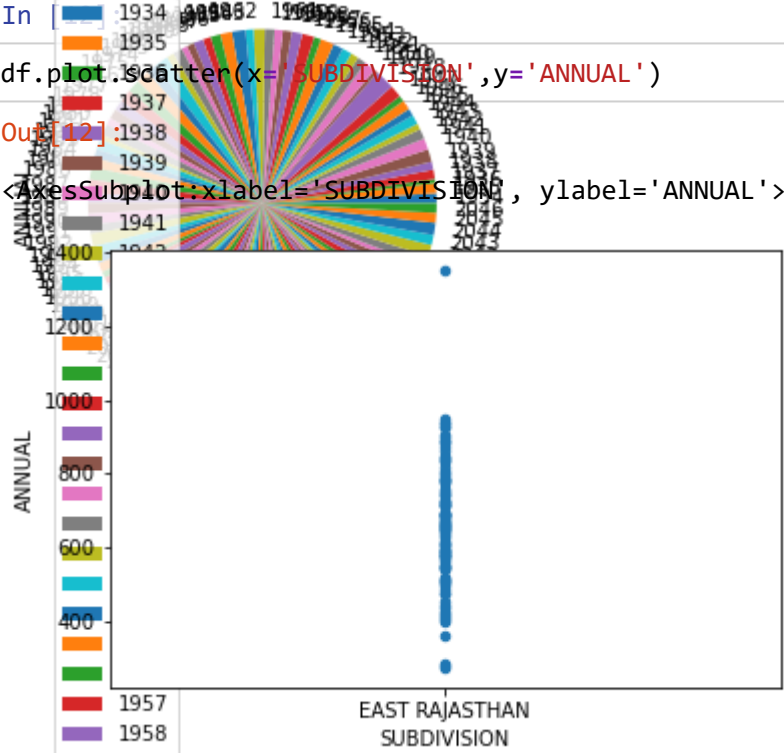
Out[11]:

<AxesSubplot:ylabel='ANNUAL'>

## Scatter chart



```
In [12]: df.plot.scatter(x='SUBDIVISION',y='ANNUAL')
Out[12]: <AxesSubplot: xlabel='SUBDIVISION', ylabel='ANNUAL'>
```

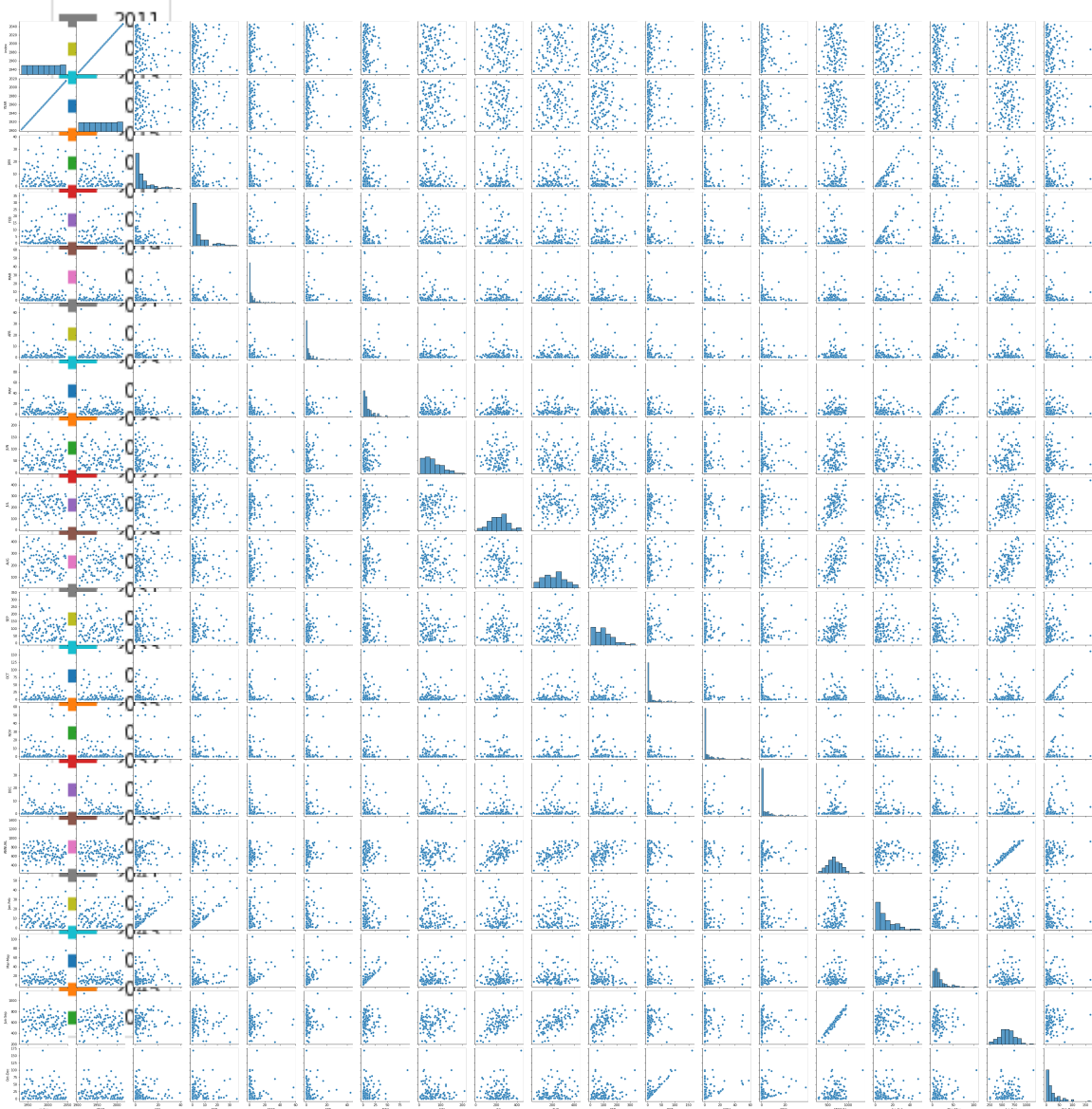


```
In [13]: df.describe()
```

	index	YEAR	JAN	FEB	MAR	APR	MAY	
count	113	113.00000	113.000000	113.000000	113.000000	113.000000	113.000000	1
mean	1959.00000	1959.00000	6.308850	5.428319	4.570796	3.177876	9.861947	
std	32.76431	32.76431	8.168303	7.516143	9.215832	5.985484	12.356991	
min	1903.00000	1903.00000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	1931.00000	1931.00000	0.700000	0.400000	0.200000	0.200000	2.300000	
50%	1959.00000	1959.00000	3.600000	2.300000	1.300000	1.100000	5.700000	
75%	1987.00000	1987.00000	8.600000	8.500000	4.100000	3.200000	12.700000	
max	2015.00000	2015.00000	39.200000	35.700000	57.400000	43.200000	90.900000	2

## EDA and Visualization

```
In [14]: sns.pairplot(df)
Out[14]: <seaborn.axisgrid.PairGrid at 0x29c5f8c9820>
```

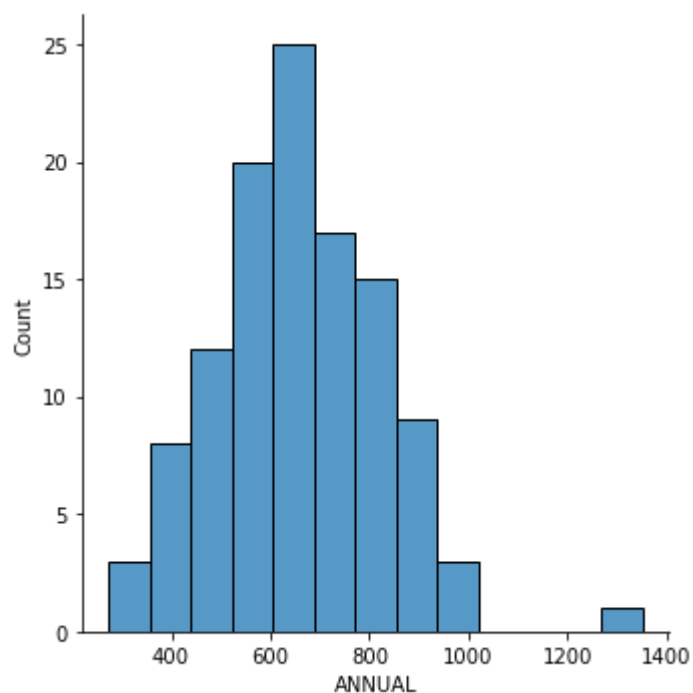


In [15]:

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

Out[15]:

```
<seaborn.axisgrid.FacetGrid at 0x29c64ce3550>
```

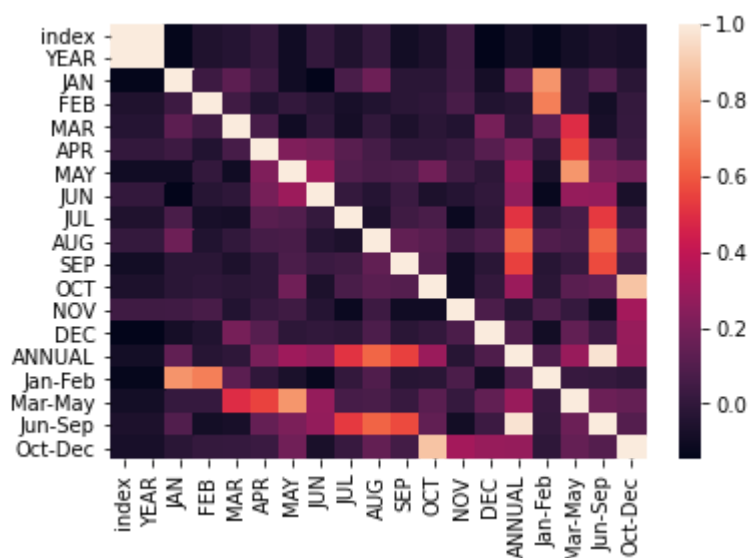


In [16]:

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

Out[16]:

```
<AxesSubplot:>
```



In [ ]:

