

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")[209:322]
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
209	209	ASSAM & MEGHALAYA	1903	19.9	25.4	103.6	140.6	206.6	607.4	362.7	551.9	306.4
210	210	ASSAM & MEGHALAYA	1904	11.1	56.1	51.9	457.1	375.2	385.7	477.6	438.8	245.9
211	211	ASSAM & MEGHALAYA	1905	19.9	16.9	137.9	213.0	275.5	521.7	439.1	649.1	276.0
212	212	ASSAM & MEGHALAYA	1906	9.7	59.2	88.4	263.5	271.3	388.4	541.8	585.9	264.2
213	213	ASSAM & MEGHALAYA	1907	55.8	37.2	109.7	254.8	213.7	509.8	578.9	304.2	401.0
...
317	317	ASSAM & MEGHALAYA	2011	11.1	11.4	109.0	92.1	238.3	316.0	395.8	302.6	221.6
318	318	ASSAM & MEGHALAYA	2012	15.2	6.9	28.8	279.1	185.8	729.7	444.3	289.2	411.6
319	319	ASSAM & MEGHALAYA	2013	1.1	9.6	44.0	112.8	346.7	286.2	367.8	289.7	229.3
320	320	ASSAM & MEGHALAYA	2014	2.0	28.3	29.3	51.5	351.1	426.4	374.4	484.6	420.2
321	321	ASSAM & MEGHALAYA	2015	13.4	15.5	37.5	250.9	332.5	558.5	300.1	590.9	279.9

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
209	209	ASSAM & MEGHALAYA	1903	19.9	25.4	103.6	140.6	206.6	607.4	362.7	551.9	306.4
210	210	ASSAM & MEGHALAYA	1904	11.1	56.1	51.9	457.1	375.2	385.7	477.6	438.8	245.9
211	211	ASSAM & MEGHALAYA	1905	19.9	16.9	137.9	213.0	275.5	521.7	439.1	649.1	276.0
212	212	ASSAM & MEGHALAYA	1906	9.7	59.2	88.4	263.5	271.3	388.4	541.8	585.9	264.2
213	213	ASSAM & MEGHALAYA	1907	55.8	37.2	109.7	254.8	213.7	509.8	578.9	304.2	401.0
...
317	317	ASSAM & MEGHALAYA	2011	11.1	11.4	109.0	92.1	238.3	316.0	395.8	302.6	221.6
318	318	ASSAM & MEGHALAYA	2012	15.2	6.9	28.8	279.1	185.8	729.7	444.3	289.2	411.6
319	319	ASSAM & MEGHALAYA	2013	1.1	9.6	44.0	112.8	346.7	286.2	367.8	289.7	229.3
320	320	ASSAM & MEGHALAYA	2014	2.0	28.3	29.3	51.5	351.1	426.4	374.4	484.6	420.2
321	321	ASSAM & MEGHALAYA	2015	13.4	15.5	37.5	250.9	332.5	558.5	300.1	590.9	279.9

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, 209 to 321
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
```

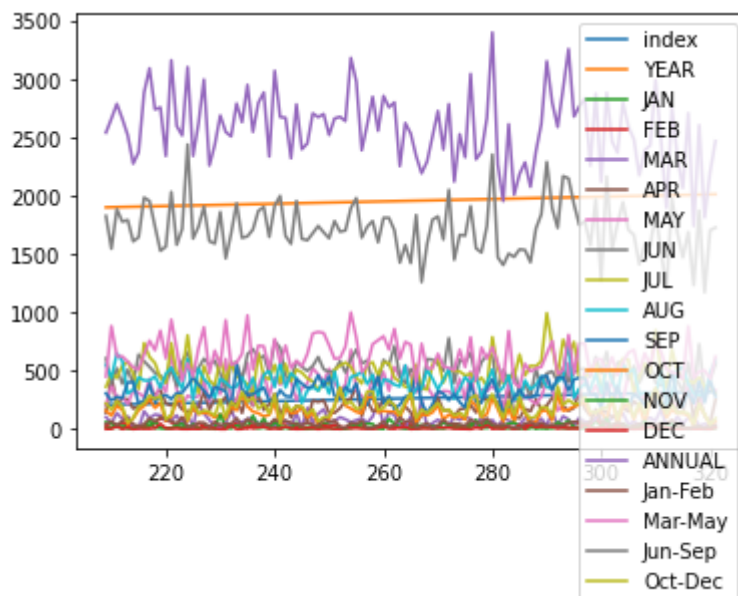
Line Chart

In [6]:

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

Out[6]:

<AxesSubplot:>



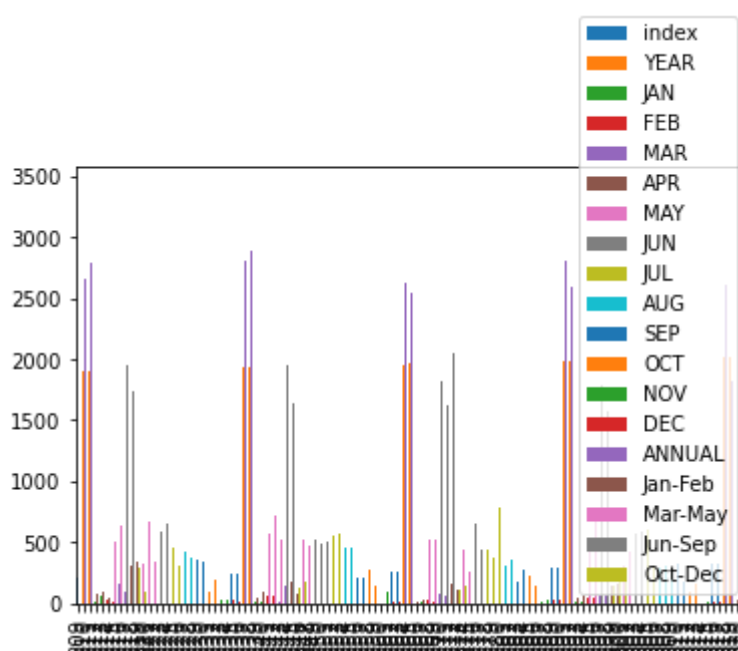
Bar chart

In [7]:

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

Out[7]:

<AxesSubplot:>



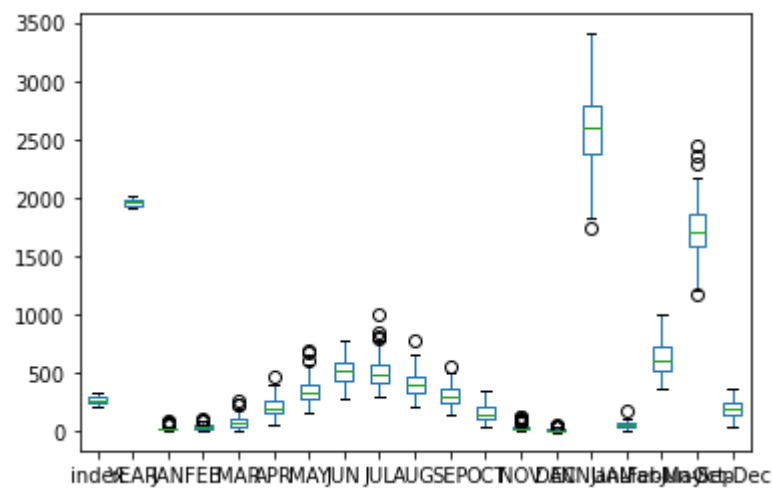
Box chart

In [8]:

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

Out[8]:

<AxesSubplot:>



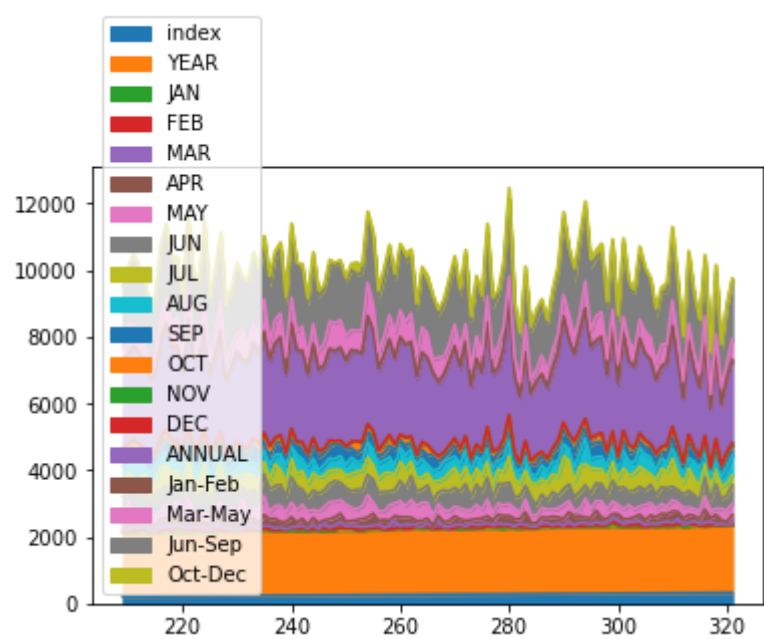
Area Chart

In [9]:

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

Out[9]:

<AxesSubplot:>



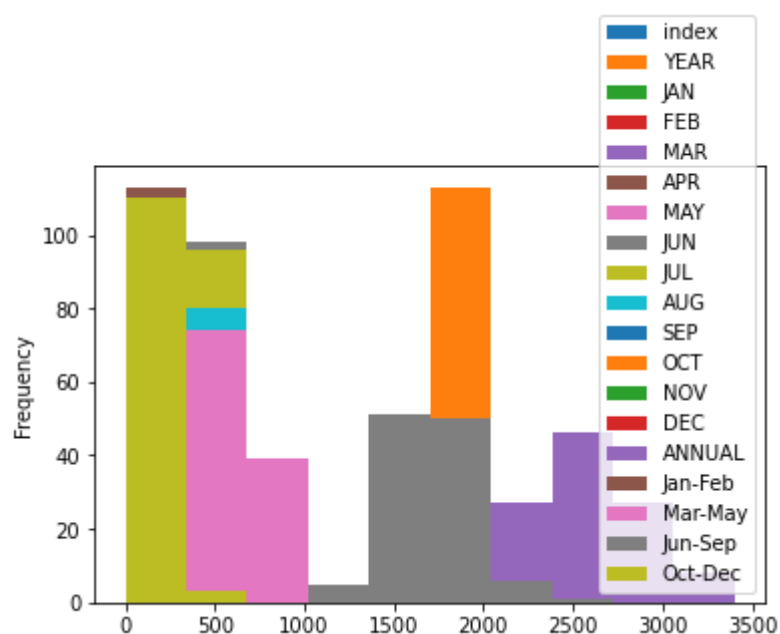
Histogram

In [10]:

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

Out[10]:

<AxesSubplot:ylabel='Frequency'>



pie chart

In [11]:

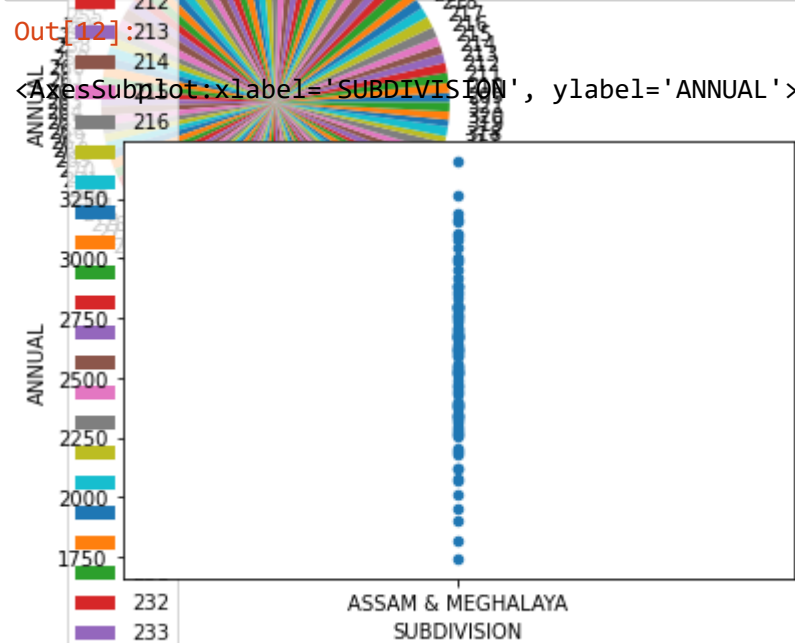
```
df.plot.pie(y='ANNUAL')
```

Out[11]:

<AxesSubplot:ylabel='ANNUAL'>

Scatter chart


```
In [12]: df.plot.scatter(x='SUBDIVISION',y='ANNUAL')
```



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

Out[13]:

	index	YEAR	JAN	FEB	MAR	APR	MAY
count	113.00000	113.00000	113.00000	113.00000	113.00000	113.00000	113.00000
mean	265.00000	1959.00000	16.953097	31.735398	79.220354	201.639823	343.432743
std	32.76431	32.76431	15.762212	21.022012	48.833248	73.664869	105.667905
min	209.00000	1903.00000	0.100000	0.500000	4.400000	45.900000	159.700000
25%	237.00000	1931.00000	7.400000	16.100000	38.700000	150.600000	266.400000
50%	265.00000	1959.00000	12.800000	29.000000	75.700000	195.000000	327.100000
75%	293.00000	1987.00000	21.200000	42.400000	103.600000	250.900000	397.800000
max	321.00000	2015.00000	83.900000	96.900000	253.800000	457.100000	681.200000

EDA and Visualization

```
In [14]: sns.pairplot(df)
```

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

In []

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

sns.heatmap(df.corr())

In []