Import Libraries

In [1]:

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

In [8]:

```
df=pd.read_csv(r"c:\Users\user\Downloads\FP2_RainFall\rainfall.csv")[0:109]
df
```

Out[8]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
0	0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6
1	1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2
2	2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0
3	3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4
4	4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0
...	...	...	...	...	...	...	...	...	...	...	...	..
104	104	ANDAMAN & NICOBAR ISLANDS	2010	101.7	8.0	0.7	12.5	319.0	448.9	521.9	563.8	263.3
105	105	ANDAMAN & NICOBAR ISLANDS	2011	265.9	84.8	272.8	111.4	326.5	383.2	583.2	441.5	757.1
106	106	ANDAMAN & NICOBAR ISLANDS	2012	119.9	45.6	30.9	55.8	533.9	458.2	317.3	369.6	868.9
107	107	ANDAMAN & NICOBAR ISLANDS	2013	67.1	37.6	43.0	46.3	509.3	777.0	564.8	336.7	473.6
108	108	ANDAMAN & NICOBAR ISLANDS	2014	41.9	8.6	0.0	11.1	238.0	416.6	467.6	321.6	412.9

109 rows × 20 columns

Data Cleaning and Preprocessing

In [9]:

```
df.dropna()
```

Out[9]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
0	0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6
1	1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2
2	2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0
3	3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4
4	4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0
...	...	...	...	...	...	...	...	...	...	...	...	..
104	104	ANDAMAN & NICOBAR ISLANDS	2010	101.7	8.0	0.7	12.5	319.0	448.9	521.9	563.8	263.3
105	105	ANDAMAN & NICOBAR ISLANDS	2011	265.9	84.8	272.8	111.4	326.5	383.2	583.2	441.5	757.1
106	106	ANDAMAN & NICOBAR ISLANDS	2012	119.9	45.6	30.9	55.8	533.9	458.2	317.3	369.6	868.9
107	107	ANDAMAN & NICOBAR ISLANDS	2013	67.1	37.6	43.0	46.3	509.3	777.0	564.8	336.7	473.6
108	108	ANDAMAN & NICOBAR ISLANDS	2014	41.9	8.6	0.0	11.1	238.0	416.6	467.6	321.6	412.9

103 rows × 20 columns

In [10]:

```
df.columns
```

Out[10]:

```
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 [11]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 109 entries, 0 to 108
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   index        109 non-null    int64  
 1   SUBDIVISION  109 non-null    object 
 2   YEAR         109 non-null    int64  
 3   JAN          109 non-null    float64
 4   FEB          109 non-null    float64
 5   MAR          107 non-null    float64
 6   APR          107 non-null    float64
 7   MAY          108 non-null    float64
 8   JUN          107 non-null    float64
 9   JUL          107 non-null    float64
 10  AUG          107 non-null    float64
 11  SEP          106 non-null    float64
 12  OCT          107 non-null    float64
 13  NOV          107 non-null    float64
 14  DEC          106 non-null    float64
 15  ANNUAL       103 non-null    float64
 16  Jan-Feb      109 non-null    float64
 17  Mar-May      106 non-null    float64
 18  Jun-Sep      106 non-null    float64
 19  Oct-Dec      106 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 17.2+ KB
```

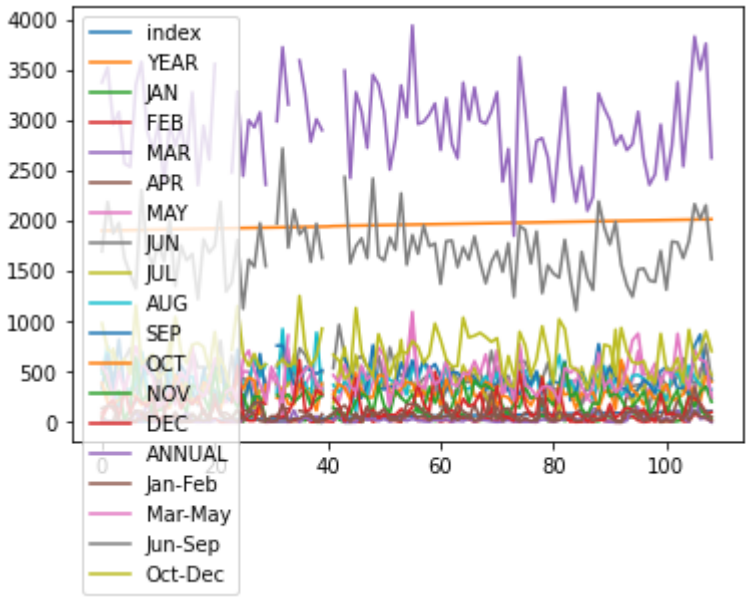
Line Chart

In [12]:

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

Out[12]:

<AxesSubplot:>



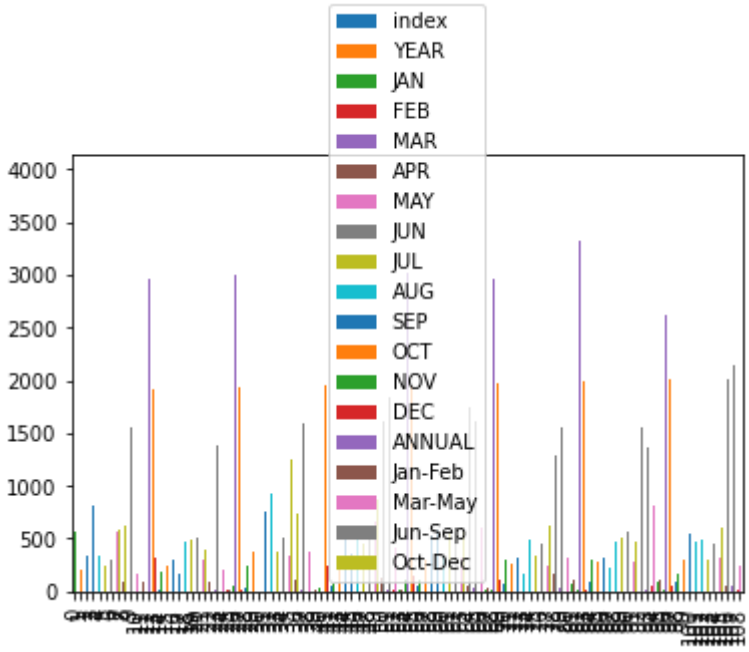
Bar chart

In [13]:

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

Out[13]:

<AxesSubplot:>



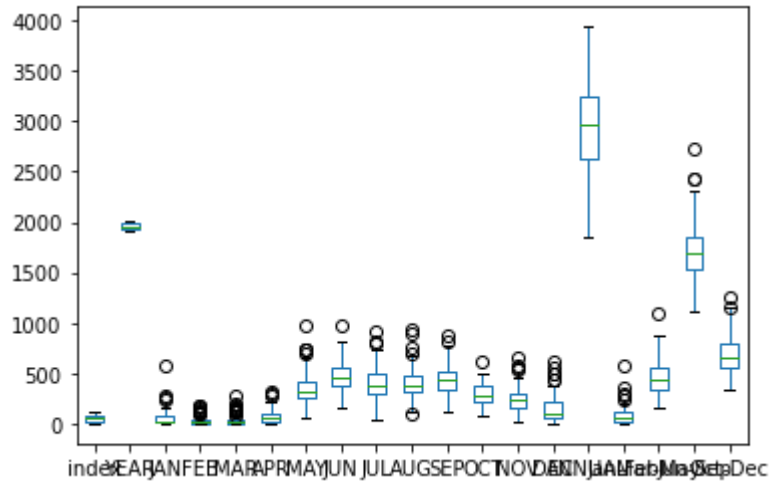
Box chart

In [14]:

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

Out[14]:

<AxesSubplot:>



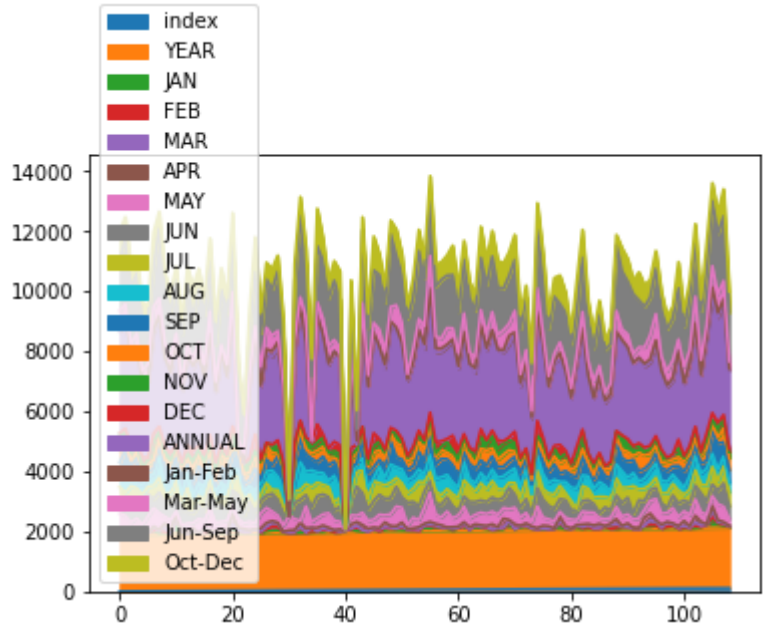
Area Chart

In [15]:

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

Out[15]:

<AxesSubplot:>



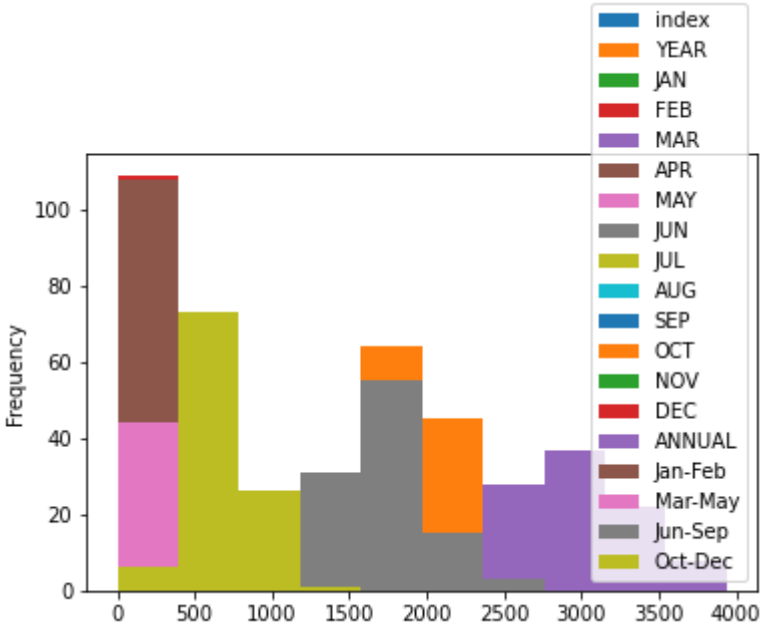
Histogram

In [17]:

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

Out[17]:

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



pie chart

In [18]:

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

Out[18]:

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

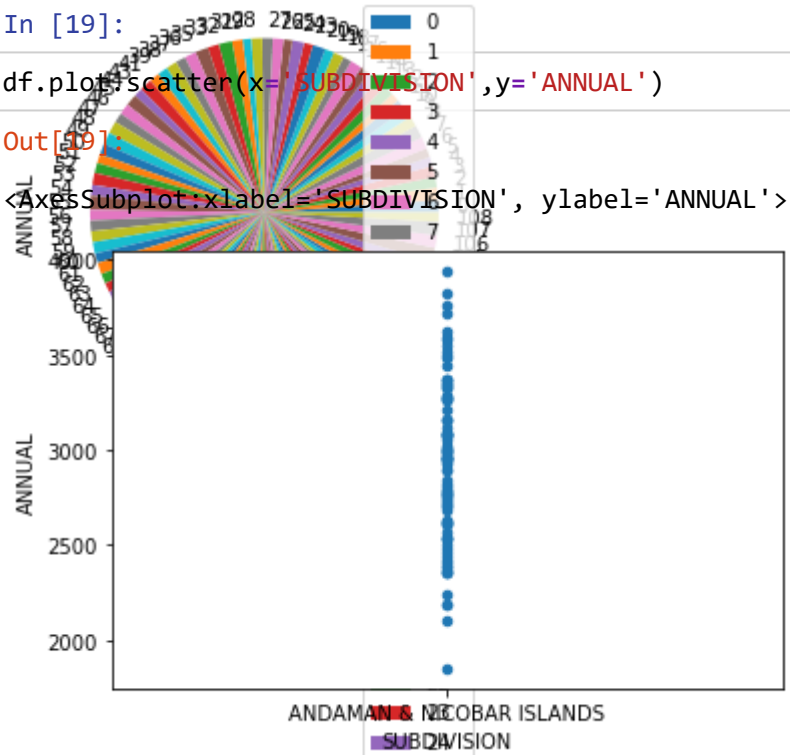
Scatter chart

In [19]:

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

Out[19]:

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



In [20]:

```
df.describe()
```

Out[20]:

	index	YEAR	JAN	FEB	MAR	APR	MAY
count	109.000000	109.000000	109.000000	109.000000	107.000000	107.000000	108.000000
mean	54.000000	1958.403670	51.956881	28.181651	32.092523	71.606542	357.289815
std	31.609598	33.384993	73.294619	38.523918	48.937113	65.927054	150.541690
min	0.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	62.000000
25%	27.000000	1929.000000	10.200000	1.700000	2.800000	21.350000	258.600000
50%	54.000000	1960.000000	28.000000	12.800000	12.200000	54.600000	318.150000
75%	81.000000	1987.000000	74.200000	35.900000	33.250000	100.900000	425.325000
max	108.000000	2014.000000	583.700000	173.800000	272.800000	323.100000	973.100000

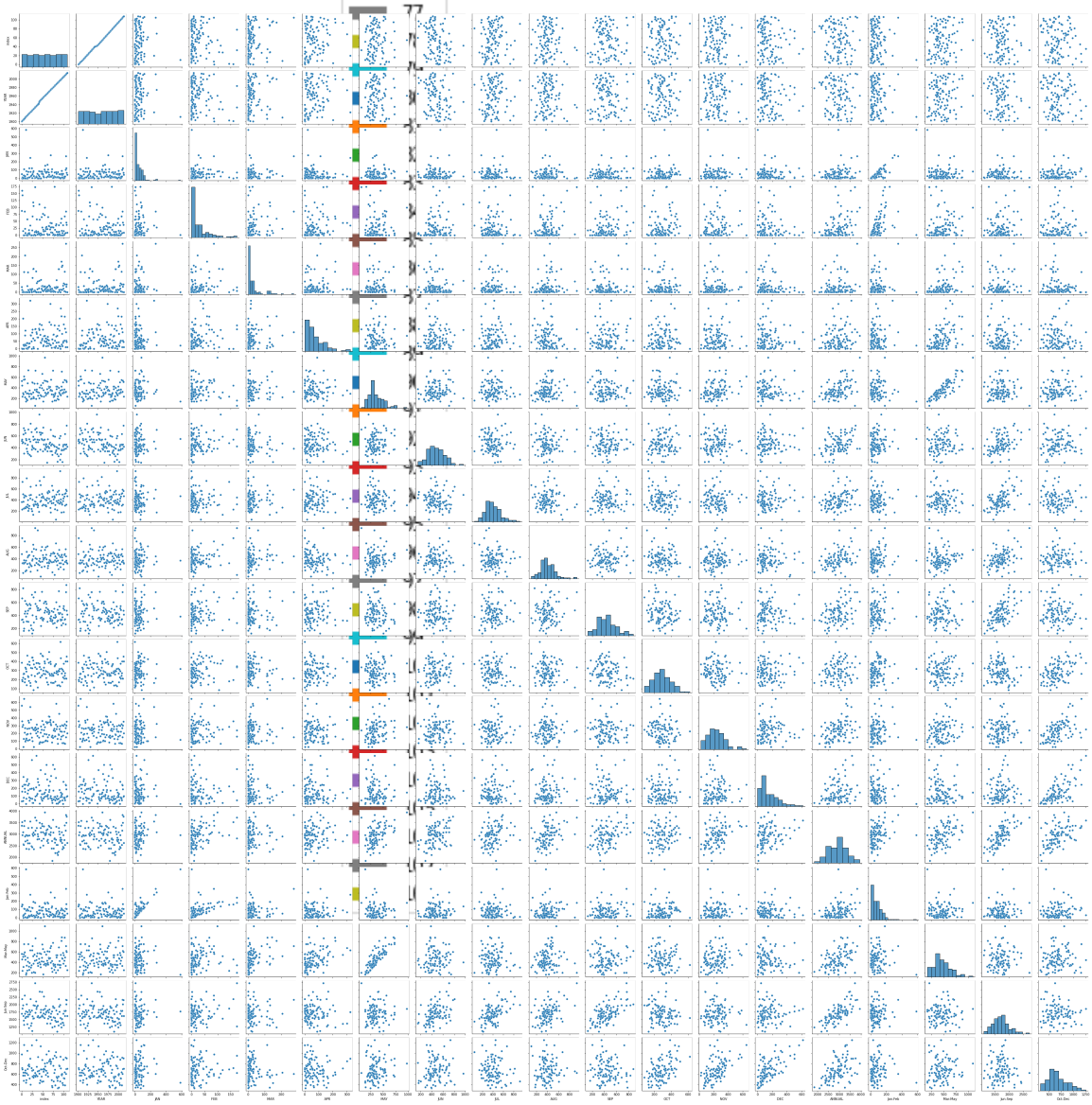
EDA and Visualization

In [21]:

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

Out[21]:

```
<seaborn.axisgrid.PairGrid at 0x1d328366790>
```

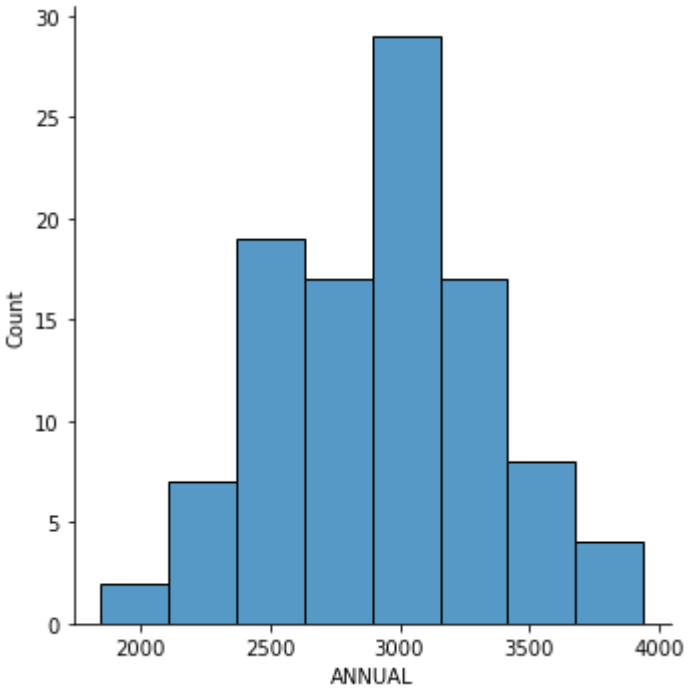


In [23]:

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

Out[23]:

```
<seaborn.axisgrid.FacetGrid at 0x1d33328da30>
```

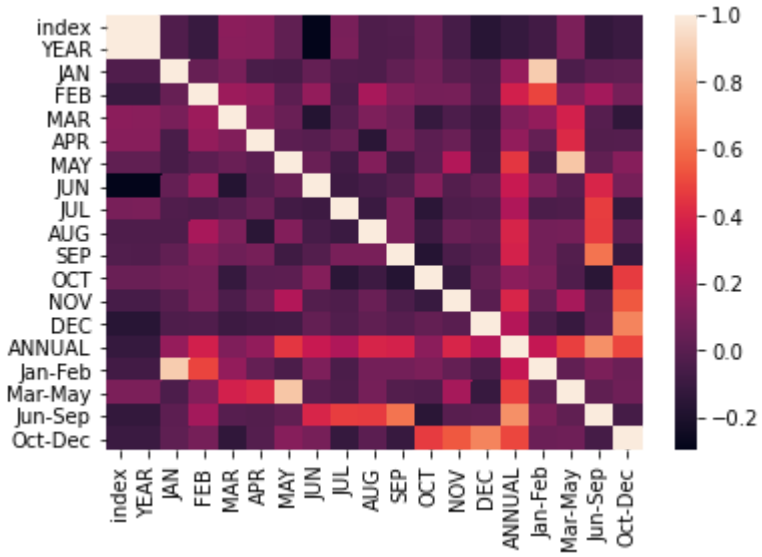


In [24]:

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

Out[24]:

```
<AxesSubplot:>
```



In [ ]: