Import Libraries

In [1]:

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

In [18]:

```python
df=pd.read_csv(r"c:\Users\user\Downloads\FP2_RainFall\rainfall.csv")[899:1012]
df
```

Out[18]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
899	899	BIHAR	1903	5.3	4.7	2.0	4.7	28.2	192.9	115.0	342.6	173.9
900	900	BIHAR	1904	6.3	1.7	3.5	5.3	118.7	191.6	394.4	351.3	84.4
901	901	BIHAR	1905	16.0	30.1	32.6	21.4	77.5	50.5	409.1	495.3	353.9
902	902	BIHAR	1906	14.4	56.9	8.6	0.9	44.7	191.3	366.2	430.1	118.7
903	903	BIHAR	1907	1.3	55.2	39.4	27.0	32.9	232.2	282.4	242.2	206.9
...	...	...	...	...	...	...	...	...	...	...	...	...
1007	1007	BIHAR	2011	4.2	7.7	9.2	23.9	74.5	211.0	241.1	278.7	234.1
1008	1008	BIHAR	2012	18.1	2.7	7.3	20.4	18.8	96.2	354.0	240.4	233.8
1009	1009	BIHAR	2013	5.1	22.6	0.6	32.3	89.5	183.3	182.0	213.6	143.3
1010	1010	BIHAR	2014	17.0	33.5	8.4	0.7	103.9	115.2	265.4	307.6	160.3
1011	1011	BIHAR	2015	12.8	1.8	27.2	38.7	39.5	122.1	231.5	287.0	101.7

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
899	899	BIHAR	1903	5.3	4.7	2.0	4.7	28.2	192.9	115.0	342.6	173.9
900	900	BIHAR	1904	6.3	1.7	3.5	5.3	118.7	191.6	394.4	351.3	84.4
901	901	BIHAR	1905	16.0	30.1	32.6	21.4	77.5	50.5	409.1	495.3	353.9
902	902	BIHAR	1906	14.4	56.9	8.6	0.9	44.7	191.3	366.2	430.1	118.7
903	903	BIHAR	1907	1.3	55.2	39.4	27.0	32.9	232.2	282.4	242.2	206.9
...	...	...	...	...	...	...	...	...	...	...	...	...
1007	1007	BIHAR	2011	4.2	7.7	9.2	23.9	74.5	211.0	241.1	278.7	234.1
1008	1008	BIHAR	2012	18.1	2.7	7.3	20.4	18.8	96.2	354.0	240.4	233.8
1009	1009	BIHAR	2013	5.1	22.6	0.6	32.3	89.5	183.3	182.0	213.6	143.3
1010	1010	BIHAR	2014	17.0	33.5	8.4	0.7	103.9	115.2	265.4	307.6	160.3
1011	1011	BIHAR	2015	12.8	1.8	27.2	38.7	39.5	122.1	231.5	287.0	101.7

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, 899 to 1011
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
```
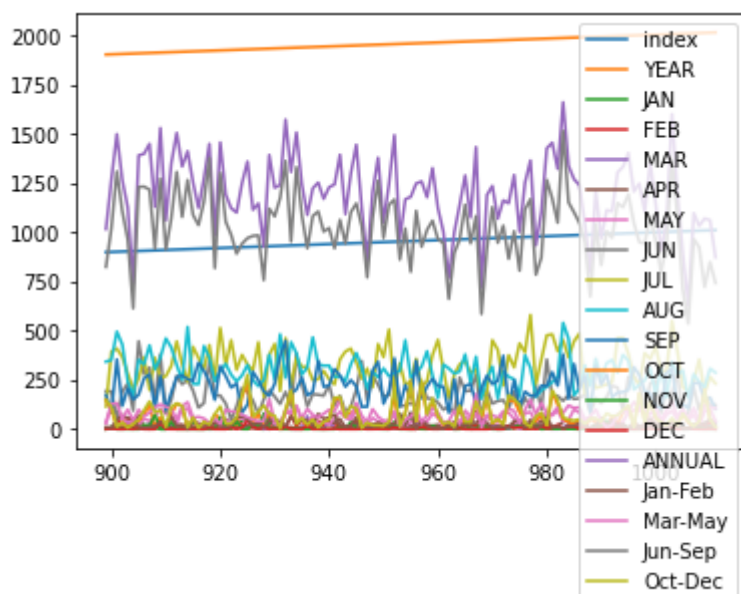
Line Chart
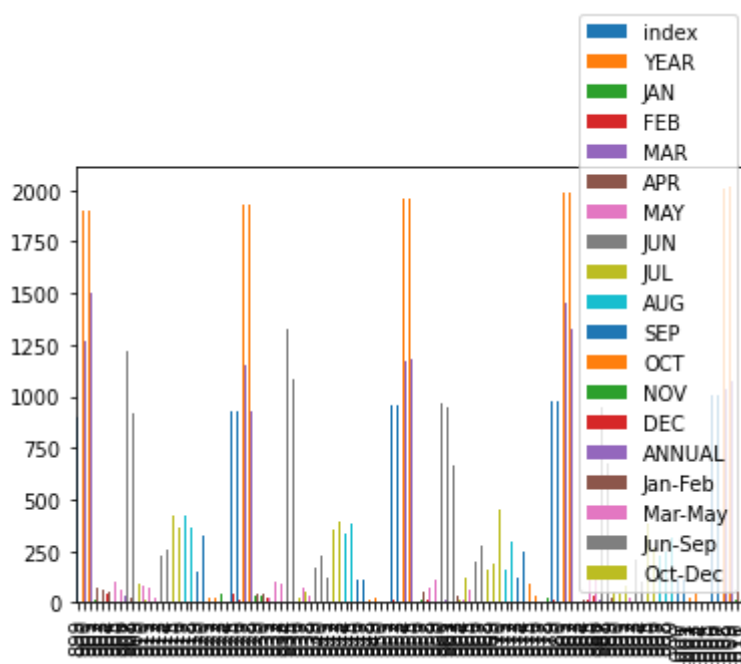
In [6]:

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

Out[6]:

<AxesSubplot:>

Bar chart

In [7]:

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

Out[7]:
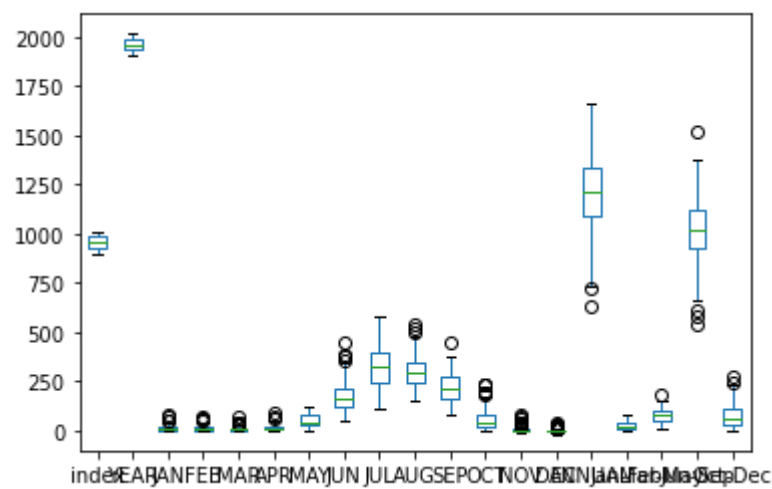
<AxesSubplot:>

Box chart

In [8]:

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

Out[8]:

`<AxesSubplot:>`
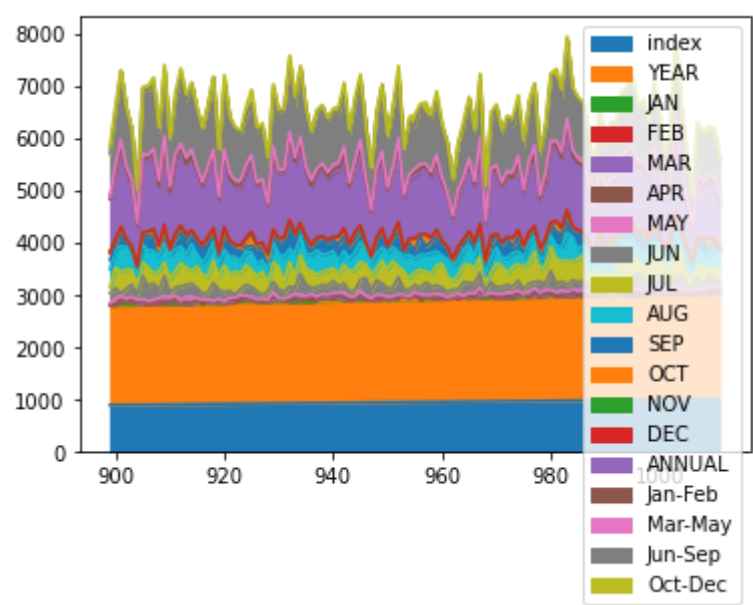

Area Chart

In [9]:

```
df.plot.area()
```
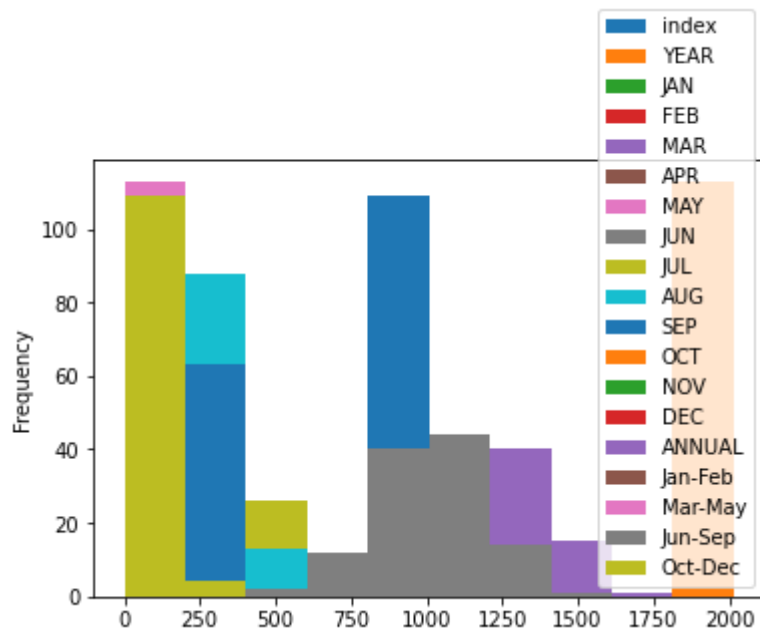
Out[9]:

`<AxesSubplot:>`

Histogram

In [10]:

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

Out[10]:

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


pie chart

In [11]:

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

Out[11]:

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

Scatter chart
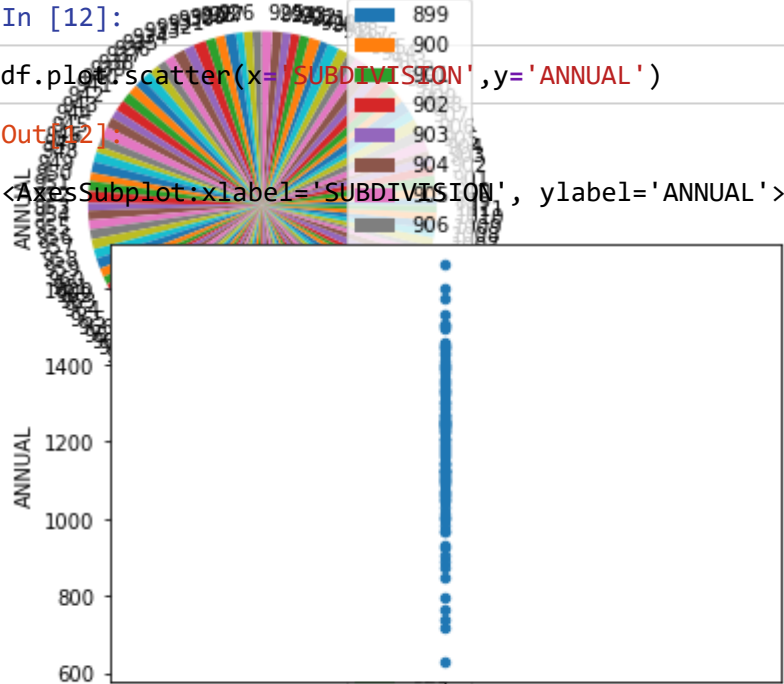
In [12]:

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

Out[12]:

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

In [13]:

```python
df.describe()
```

Out[13]:

	index	YEAR	JAN	FEB	MAR	APR	MAY
count	113.00000	113.00000	113.000000	113.000000	113.000000	113.000000	113.000000	1
mean	955.00000	1959.00000	13.123894	14.469027	9.983186	17.054867	52.853982	1
std	32.76431	32.76431	15.449062	15.145721	11.720966	16.050239	28.136625	
min	899.00000	1903.00000	0.000000	0.000000	0.000000	0.100000	1.300000	
25%	927.00000	1931.00000	2.300000	2.800000	1.800000	5.300000	31.200000	1
50%	955.00000	1959.00000	9.400000	8.400000	6.100000	12.600000	44.700000	1
75%	983.00000	1987.00000	18.100000	21.500000	12.100000	24.500000	77.000000	2
max	1011.00000	2015.00000	81.200000	66.300000	65.500000	91.400000	118.700000	4

EDA and Visualization
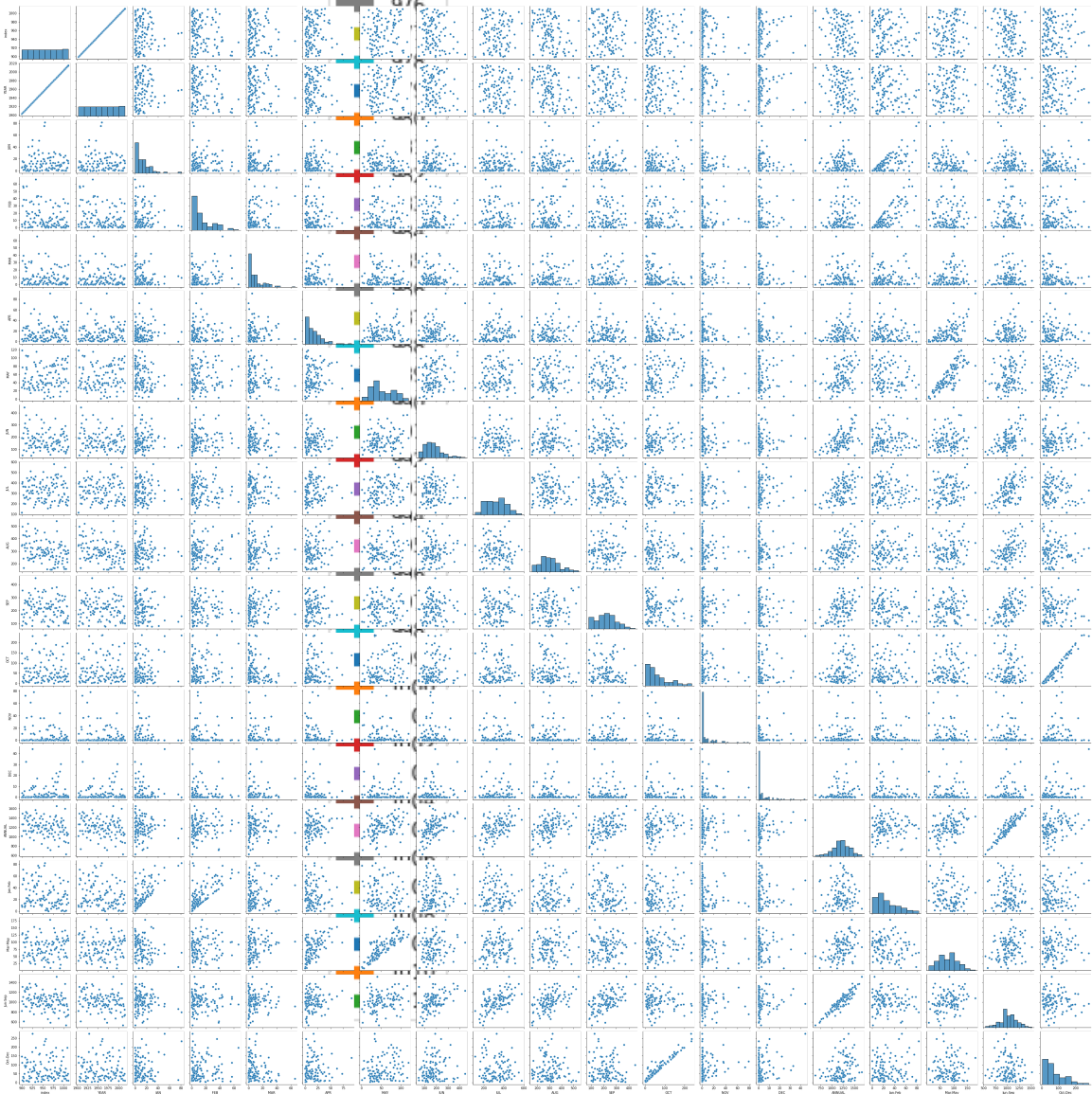
In [14]:

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

Out[14]:

<seaborn.axisgrid.PairGrid at 0x23e217739d0>

In [15]:

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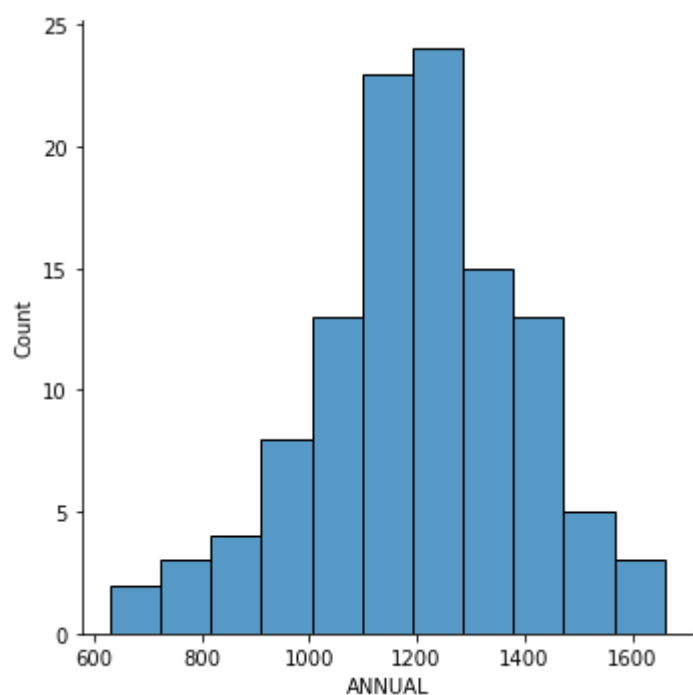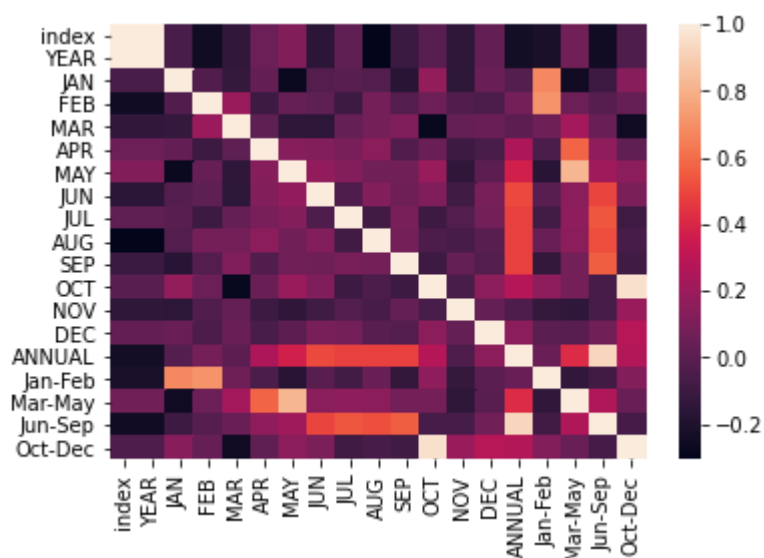
Out[15]:

```
<seaborn.axisgrid.FacetGrid at 0x23e2fde4430>
```


In [16]:

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

Out[16]:

```
<AxesSubplot:>
```


In [ ]: