Import Libraries

In [1]:

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

In [2]:

```
d=pd.read_csv(r"C:\Users\user\Downloads\FP2_RainFall\rain.csv")[669:782]
d
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
3659	3659	NORTH INTERIOR KARNATAKA	1903	3.5	0.0	0.1	6.9	53.4	102.8	209.4	146.4	189.3
3660	3660	NORTH INTERIOR KARNATAKA	1904	0.2	0.3	8.5	11.0	46.3	120.6	91.6	48.5	165.1
3661	3661	NORTH INTERIOR KARNATAKA	1905	0.0	6.0	2.6	16.0	51.2	99.6	60.1	139.2	42.2
3662	3662	NORTH INTERIOR KARNATAKA	1906	21.3	0.0	0.2	2.6	30.0	142.0	120.3	182.1	116.0
3663	3663	NORTH INTERIOR KARNATAKA	1907	1.1	0.0	12.1	96.9	5.6	90.6	155.1	185.2	165.8
...	...	...	...	...	...	...	...	...	...	...	...	...
3767	3767	NORTH INTERIOR KARNATAKA	2011	0.5	7.2	7.2	41.2	46.8	101.3	150.8	152.0	69.0
3768	3768	NORTH INTERIOR KARNATAKA	2012	28.5	6.2	0.4	35.4	19.5	60.0	114.5	105.5	79.2
3769	3769	NORTH INTERIOR KARNATAKA	2013	1.2	6.1	3.0	25.4	47.4	99.4	160.7	73.9	201.0
3770	3770	NORTH INTERIOR KARNATAKA	2014	0.0	6.1	29.2	26.4	93.0	50.4	136.8	205.2	90.2
3771	3771	NORTH INTERIOR KARNATAKA	2015	2.4	0.0	27.5	50.8	45.3	89.6	38.5	78.4	150.8

113 rows × 20 columns

Data Cleaning and preprocessing

In [3]:

```
d.dropna()
```

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
3659	3659	NORTH INTERIOR KARNATAKA	1903	3.5	0.0	0.1	6.9	53.4	102.8	209.4	146.4	189.3
3660	3660	NORTH INTERIOR KARNATAKA	1904	0.2	0.3	8.5	11.0	46.3	120.6	91.6	48.5	165.1
3661	3661	NORTH INTERIOR KARNATAKA	1905	0.0	6.0	2.6	16.0	51.2	99.6	60.1	139.2	42.2
3662	3662	NORTH INTERIOR KARNATAKA	1906	21.3	0.0	0.2	2.6	30.0	142.0	120.3	182.1	116.0
3663	3663	NORTH INTERIOR KARNATAKA	1907	1.1	0.0	12.1	96.9	5.6	90.6	155.1	185.2	165.8
...	...	...	...	...	...	...	...	...	...	...	...	...
3767	3767	NORTH INTERIOR KARNATAKA	2011	0.5	7.2	7.2	41.2	46.8	101.3	150.8	152.0	69.0
3768	3768	NORTH INTERIOR KARNATAKA	2012	28.5	6.2	0.4	35.4	19.5	60.0	114.5	105.5	79.2
3769	3769	NORTH INTERIOR KARNATAKA	2013	1.2	6.1	3.0	25.4	47.4	99.4	160.7	73.9	201.0
3770	3770	NORTH INTERIOR KARNATAKA	2014	0.0	6.1	29.2	26.4	93.0	50.4	136.8	205.2	90.2
3771	3771	NORTH INTERIOR KARNATAKA	2015	2.4	0.0	27.5	50.8	45.3	89.6	38.5	78.4	150.8

113 rows × 20 columns

In [4]:

```
d.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]:

```
d.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113 entries, 3659 to 3771
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
```

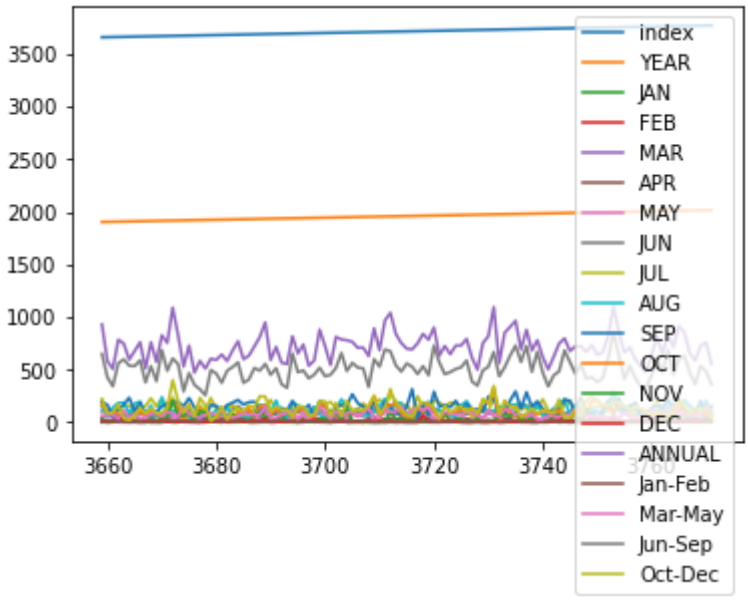
Line Chart

In [6]:

```
d.plot.line()
```

Out[6]:

<AxesSubplot:>



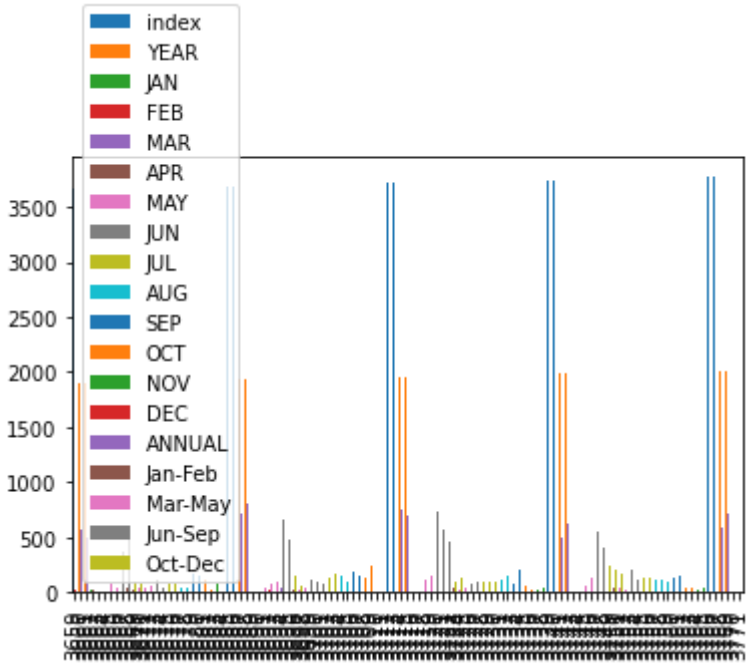
Bar Chart

In [7]:

```
d.plot.bar()
```

Out[7]:

<AxesSubplot:>



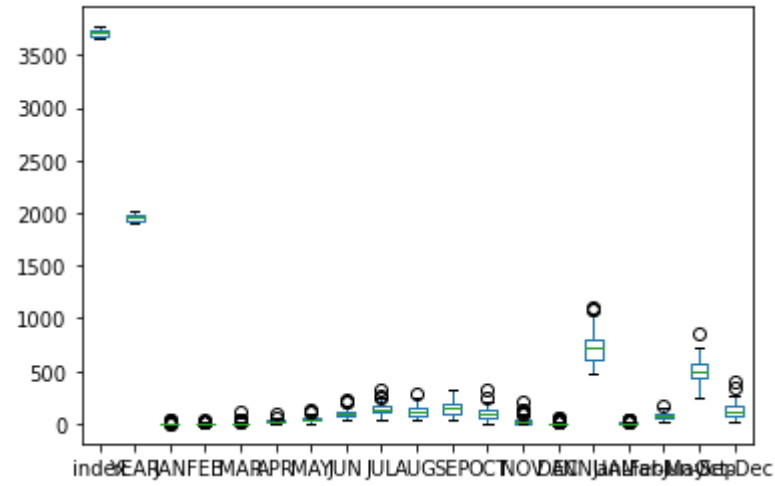
Box Chart

In [8]:

```
d.plot.box()
```

Out[8]:

<AxesSubplot:>



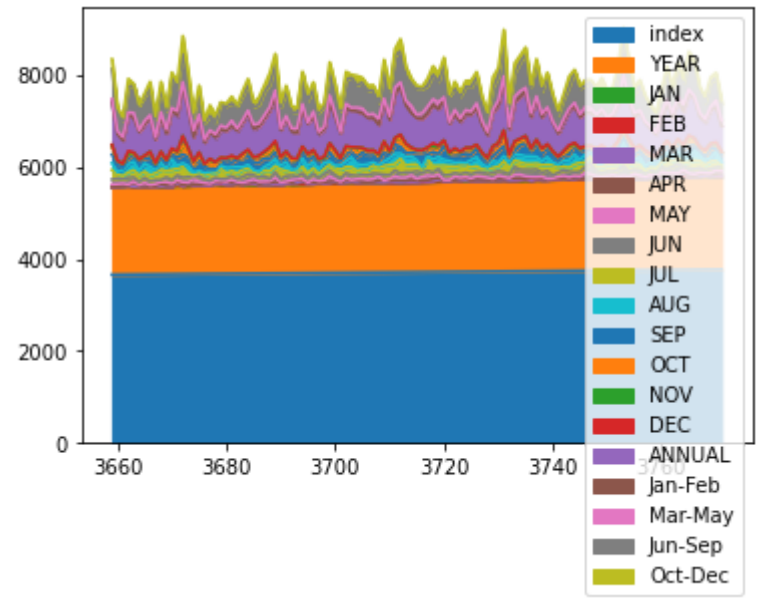
Area Chart

In [9]:

```
d.plot.area()
```

Out[9]:

<AxesSubplot:>



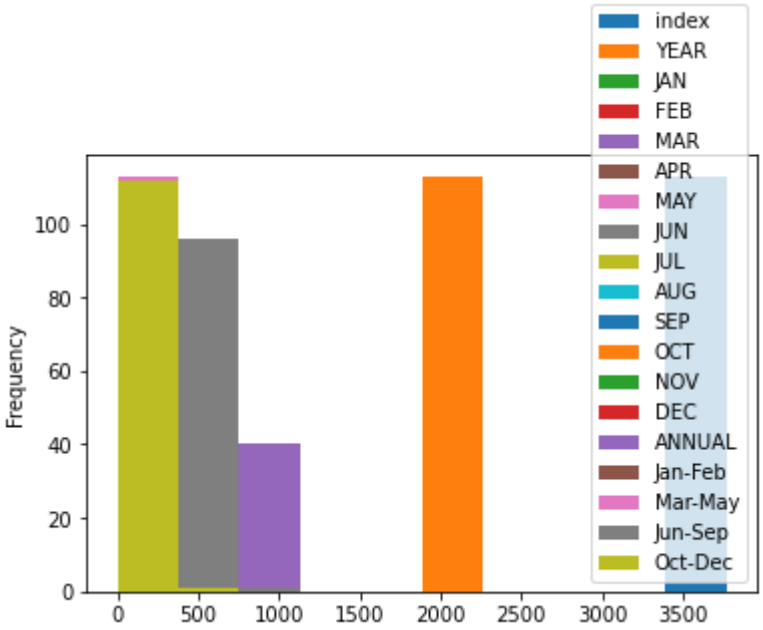
Histogram

In [10]:

```
d.plot.hist()
```

Out[10]:

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



Pie Chart

In [11]:

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

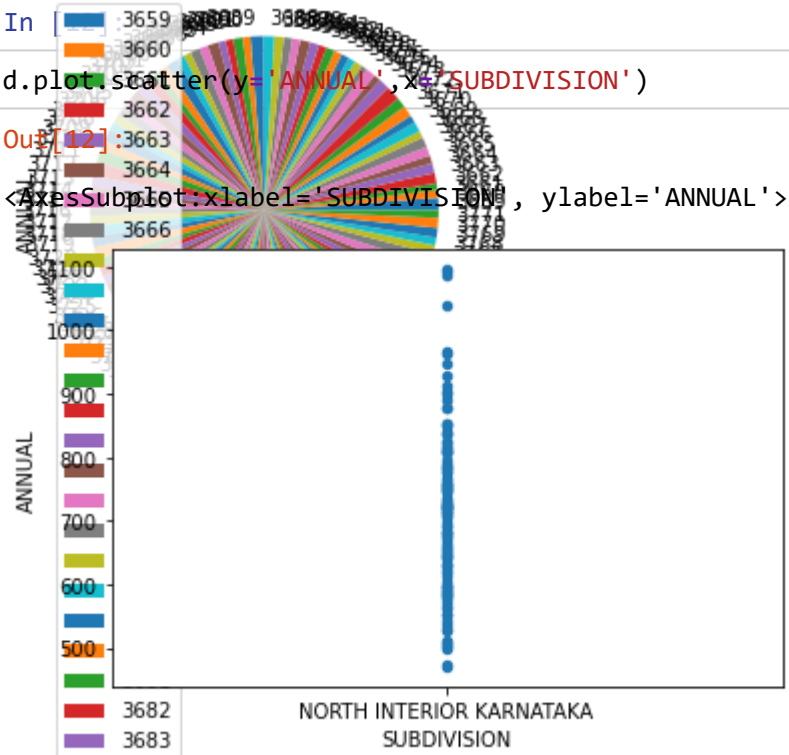
Out[11]:

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

Scatter Chart

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

Out[12]:
```
<AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>
```



In [ ]:
```
d.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	3715.00000	1959.00000	3.035398	3.061947	7.184071	23.937168	46.984071	1
std	32.76431	32.76431	6.246055	5.376741	12.767576	15.304004	26.729119	
min	3659.00000	1903.00000	0.000000	0.000000	0.000000	0.200000	3.500000	
25%	3687.00000	1931.00000	0.000000	0.000000	0.500000	12.100000	29.000000	
50%	3715.00000	1959.00000	0.200000	0.300000	3.800000	22.100000	40.600000	
75%	3743.00000	1987.00000	2.400000	3.400000	7.700000	32.100000	63.100000	1
max	3771.00000	2015.00000	28.500000	28.400000	109.200000	96.900000	127.300000	2

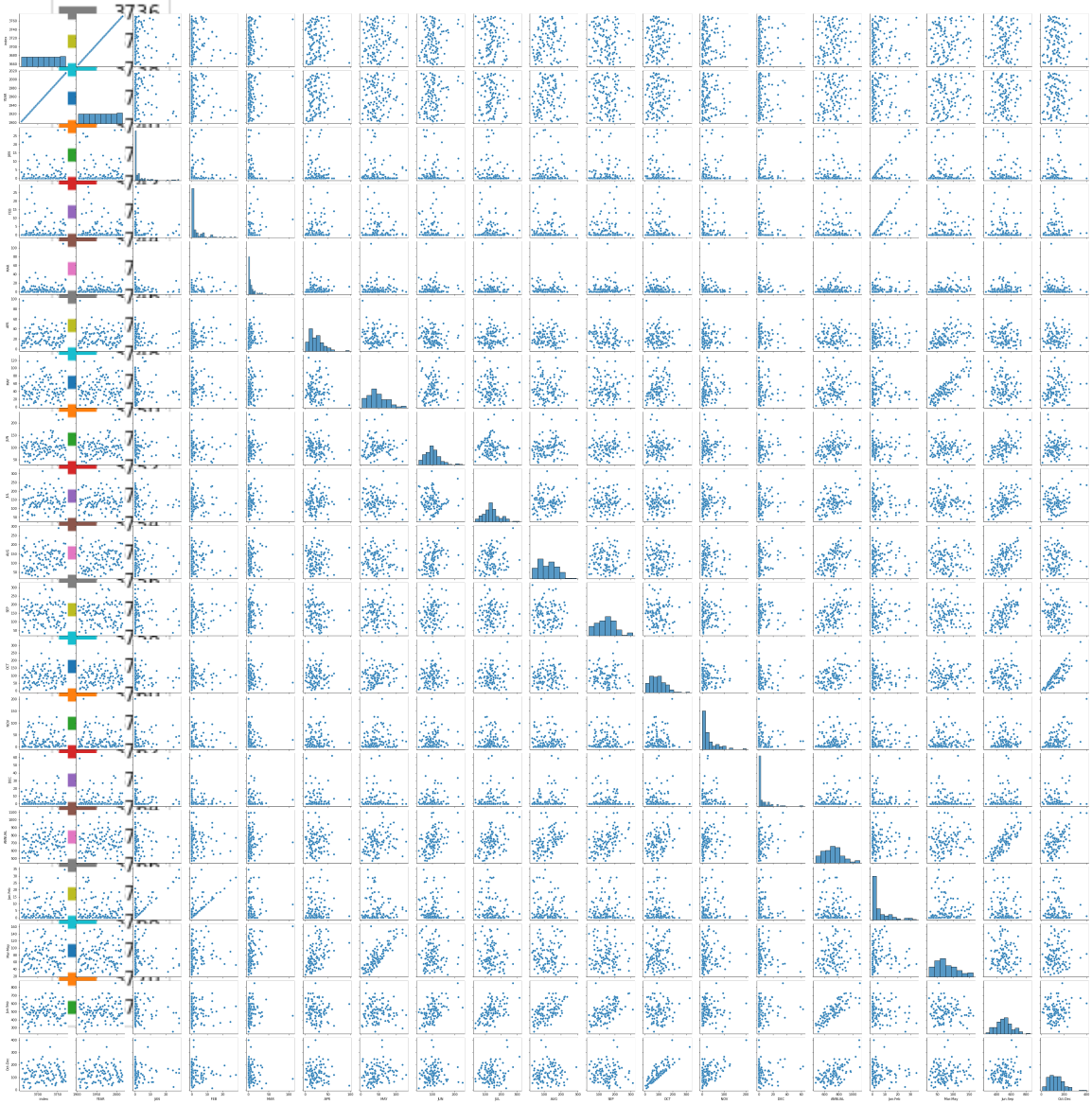
EDA AND VISUALIZATION

In [14]:

```
sns.pairplot(d)
```

Out[14]:

```
<seaborn.axisgrid.PairGrid at 0x13ae7633e80>
```

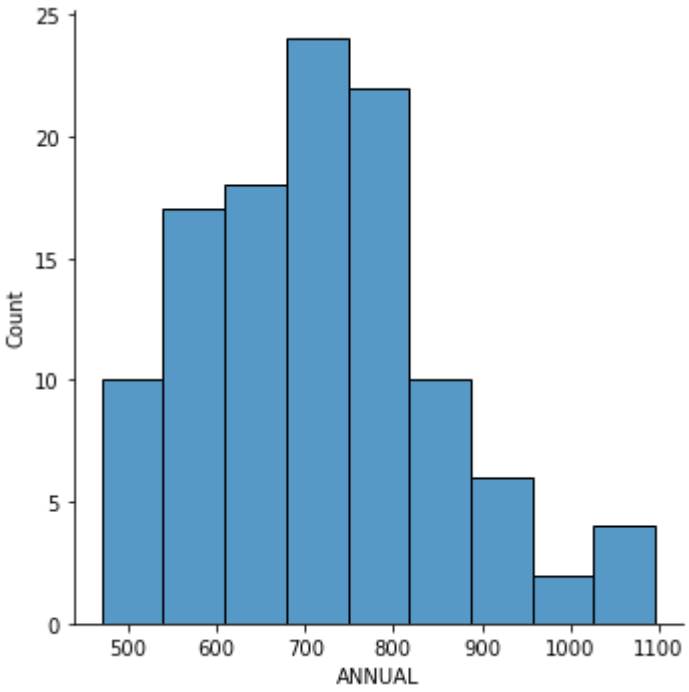


In [15]:

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

Out[15]:

```
<seaborn.axisgrid.FacetGrid at 0x13af5be0610>
```

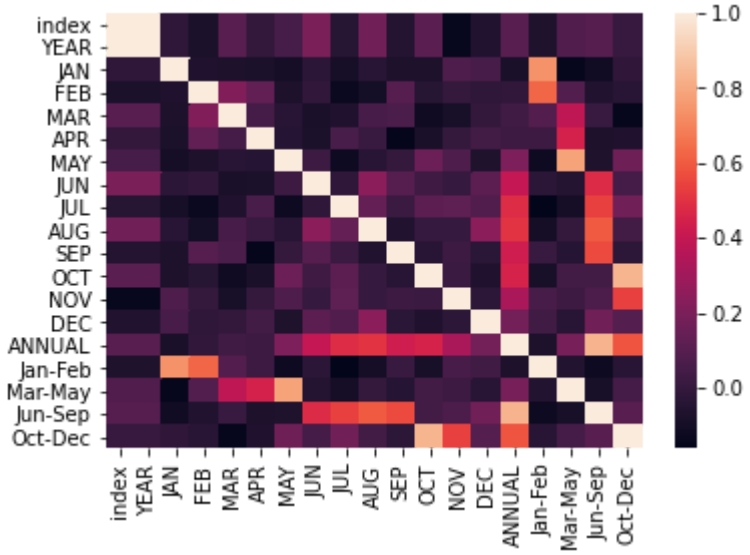


In [16]:

```
sns.heatmap(d.corr())
```

Out[16]:

```
<AxesSubplot:>
```



In [ ]: