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
669	669	ORISSA	1903	19.7	18.9	10.5	34.6	73.3	154.3	410.4	295.2	265.6
670	670	ORISSA	1904	0.2	12.2	20.6	10.1	100.2	342.9	336.7	350.4	227.8
671	671	ORISSA	1905	24.3	17.2	66.3	56.9	107.5	92.0	330.1	281.4	344.1
672	672	ORISSA	1906	19.8	88.3	40.9	5.0	46.3	219.0	348.8	261.8	251.7
673	673	ORISSA	1907	0.1	16.9	57.4	105.1	39.8	226.3	203.9	531.8	204.8
...	...	...	...	...	...	...	...	...	...	...	...	...
777	777	ORISSA	2011	3.7	16.2	4.9	58.2	75.6	210.1	199.6	358.6	398.7
778	778	ORISSA	2012	50.8	3.6	0.9	34.8	21.3	169.6	324.3	417.0	242.4
779	779	ORISSA	2013	3.3	7.8	2.1	53.6	57.7	272.6	380.0	254.9	208.1
780	780	ORISSA	2014	0.0	17.6	25.1	11.7	111.9	92.2	496.2	386.3	281.1
781	781	ORISSA	2015	15.1	3.3	10.5	67.6	32.6	238.6	294.8	264.0	237.0

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
669	669	ORISSA	1903	19.7	18.9	10.5	34.6	73.3	154.3	410.4	295.2	265.6
670	670	ORISSA	1904	0.2	12.2	20.6	10.1	100.2	342.9	336.7	350.4	227.8
671	671	ORISSA	1905	24.3	17.2	66.3	56.9	107.5	92.0	330.1	281.4	344.1
672	672	ORISSA	1906	19.8	88.3	40.9	5.0	46.3	219.0	348.8	261.8	251.7
673	673	ORISSA	1907	0.1	16.9	57.4	105.1	39.8	226.3	203.9	531.8	204.8
...	...	...	...	...	...	...	...	...	...	...	...	...
777	777	ORISSA	2011	3.7	16.2	4.9	58.2	75.6	210.1	199.6	358.6	398.7
778	778	ORISSA	2012	50.8	3.6	0.9	34.8	21.3	169.6	324.3	417.0	242.4
779	779	ORISSA	2013	3.3	7.8	2.1	53.6	57.7	272.6	380.0	254.9	208.1
780	780	ORISSA	2014	0.0	17.6	25.1	11.7	111.9	92.2	496.2	386.3	281.1
781	781	ORISSA	2015	15.1	3.3	10.5	67.6	32.6	238.6	294.8	264.0	237.0

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, 669 to 781
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
```

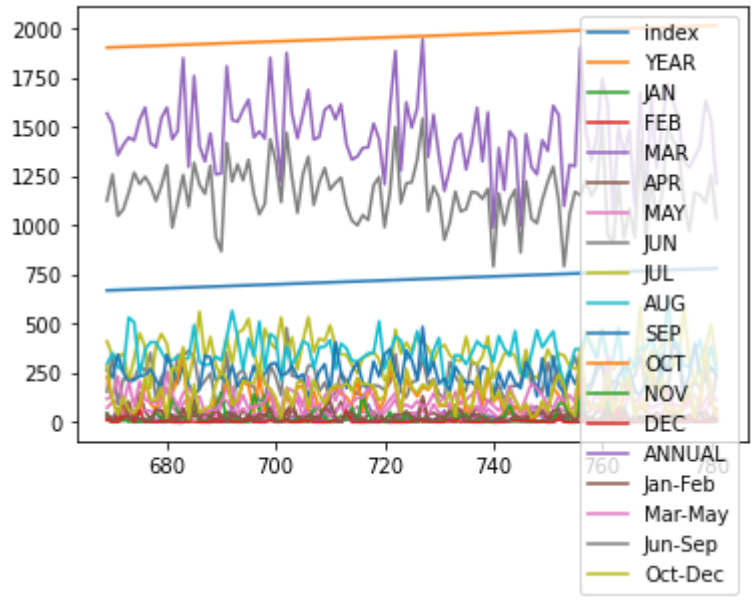
Line Chart

In [6]:

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

Out[6]:

```
<AxesSubplot:>
```



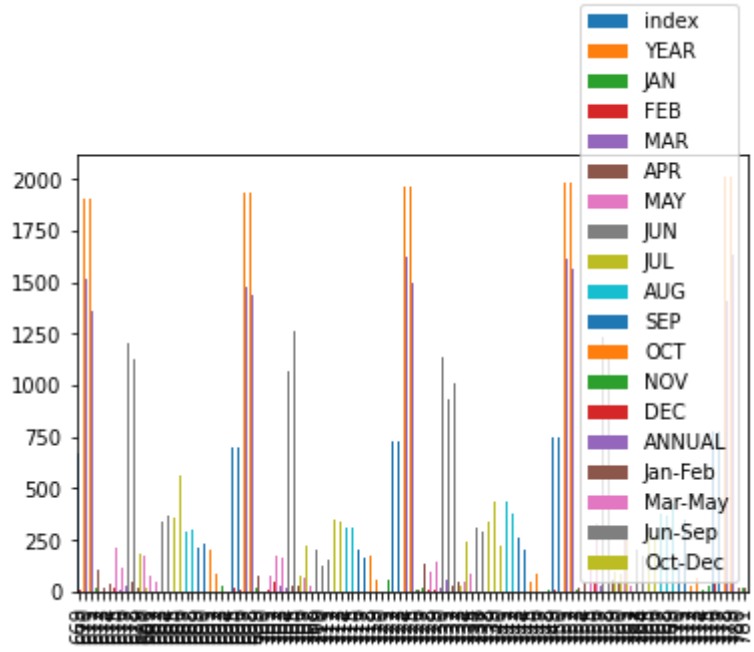
Bar Chart

In [7]:

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

Out[7]:

```
<AxesSubplot:>
```



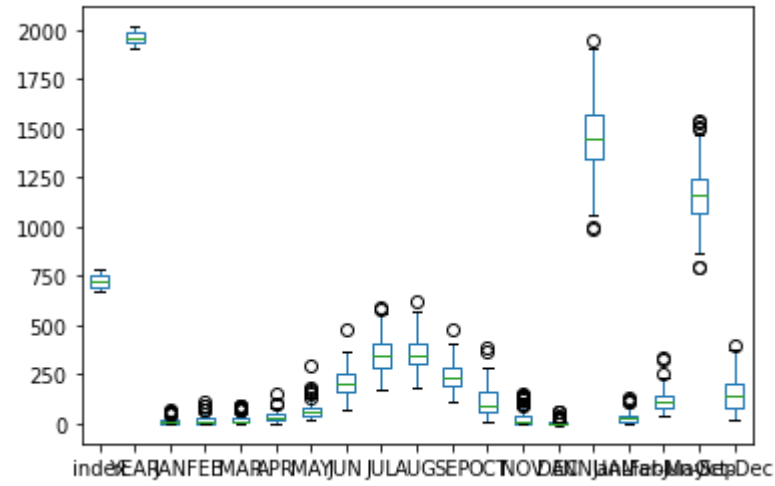
Box Chart

In [8]:

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

Out[8]:

<AxesSubplot:>



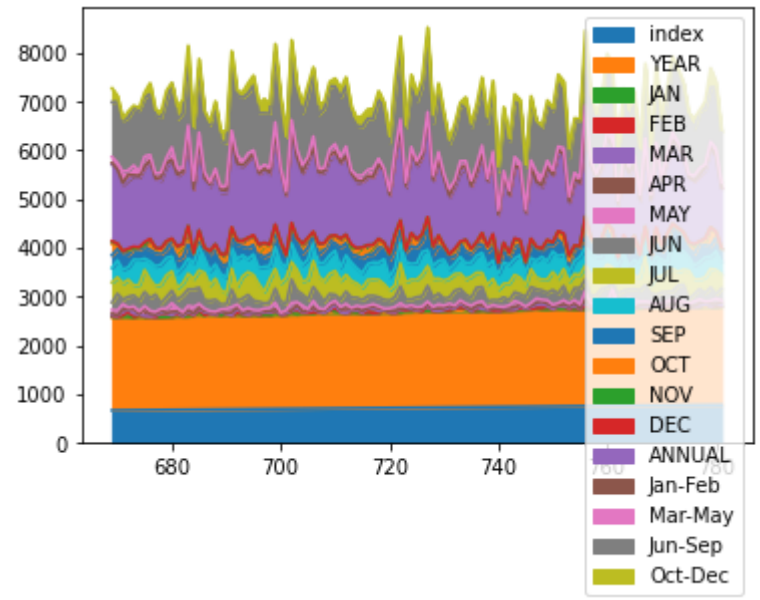
Area Chart

In [9]:

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

Out[9]:

<AxesSubplot:>



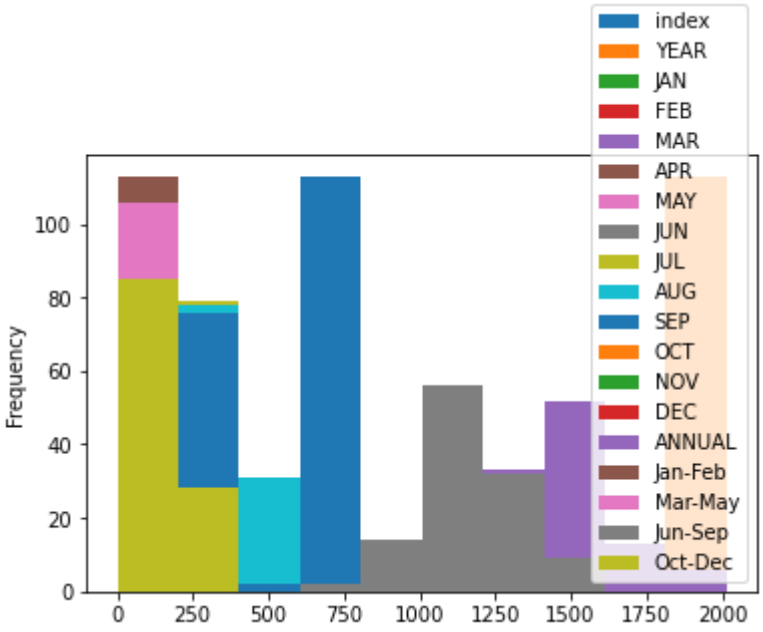
Histogram

In [10]:

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

Out[10]:

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



Pie Chart

In [11]:

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

Out[11]:

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

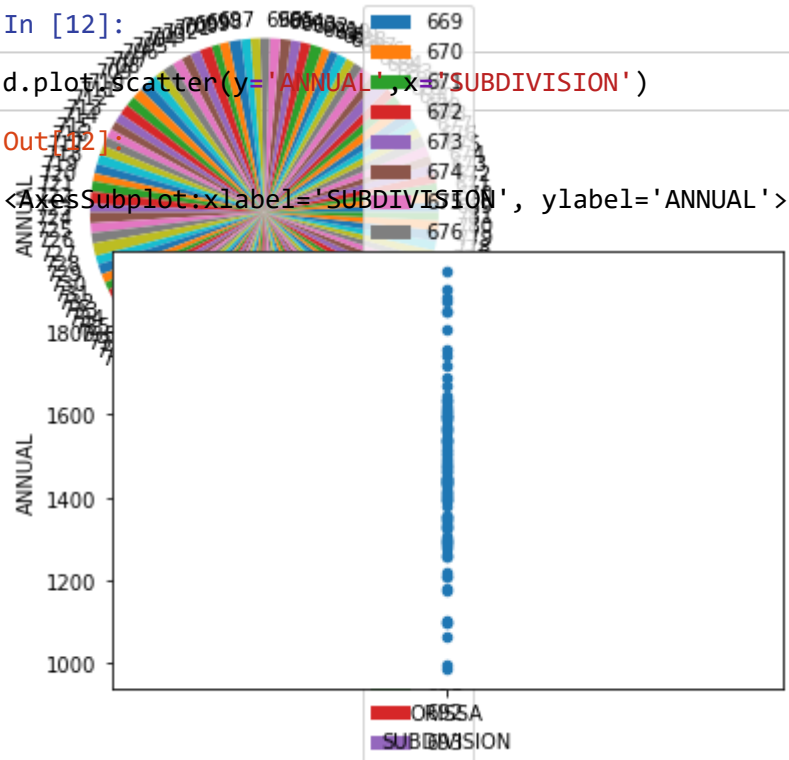
Scatter Chart

In [12]:

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

Out[12]:

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



In [13]:

```
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	11
mean	725.00000	1959.00000	12.168142	19.490265	21.240708	33.413274	64.833628	21
std	32.76431	32.76431	15.525413	21.942685	22.442552	22.508155	40.175660	7
min	669.00000	1903.00000	0.000000	0.000000	0.100000	1.100000	16.200000	7
25%	697.00000	1931.00000	0.400000	3.600000	5.400000	18.200000	38.200000	15
50%	725.00000	1959.00000	5.500000	12.100000	12.400000	28.200000	57.700000	19
75%	753.00000	1987.00000	18.000000	28.300000	27.400000	46.000000	76.500000	25
max	781.00000	2015.00000	70.900000	116.200000	89.800000	148.400000	293.000000	47

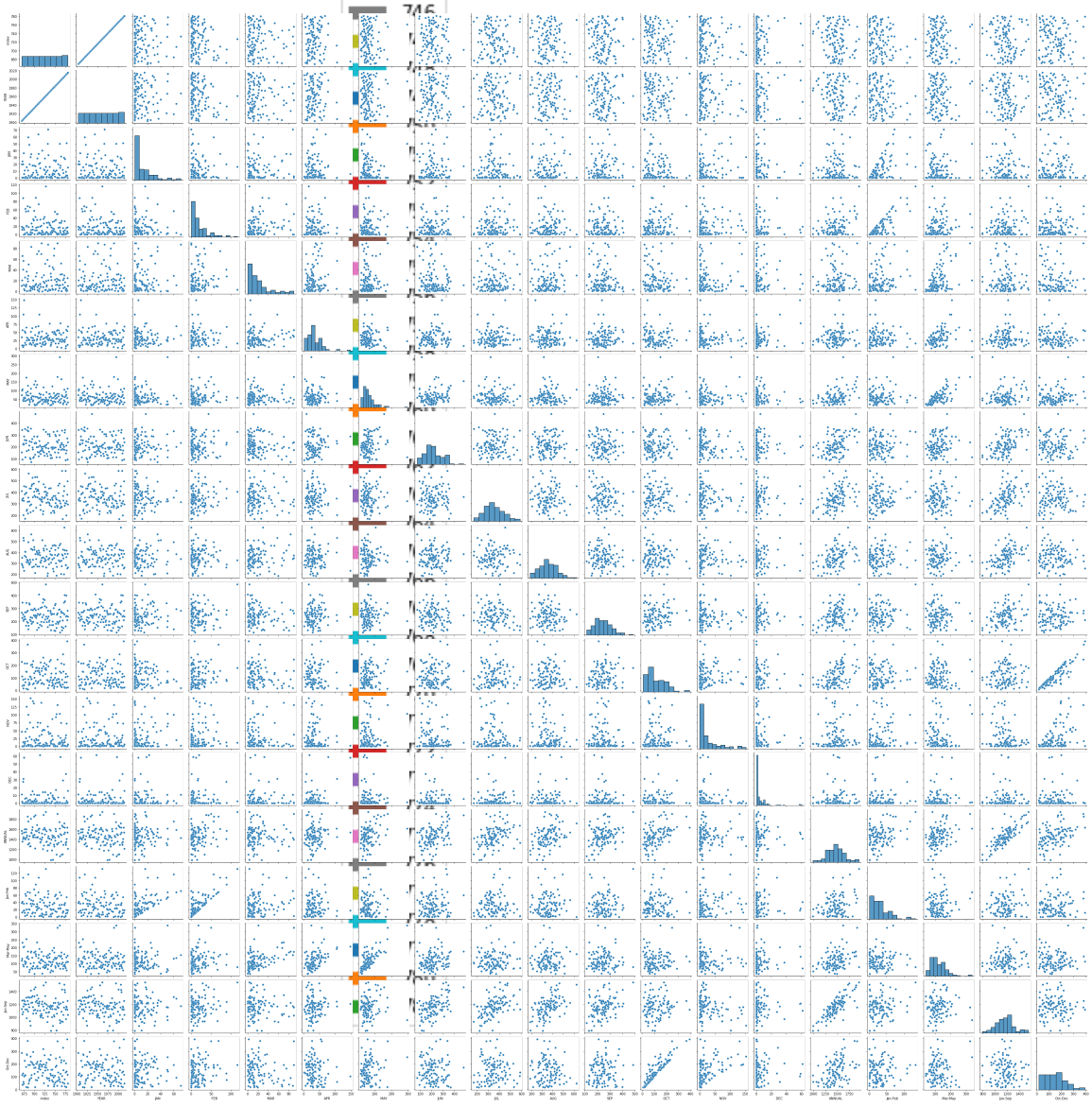
EDA AND VISUALIZATION

In [14]:

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

Out[14]:

```
<seaborn.axisgrid.PairGrid at 0x1f00cf38a00>
```

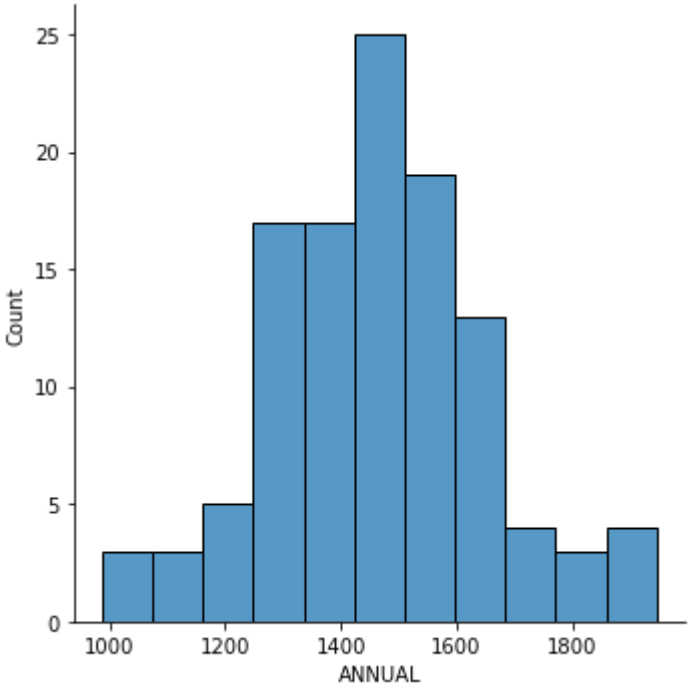


In [15]:

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

Out[15]:

```
<seaborn.axisgrid.FacetGrid at 0x1f0193b3430>
```

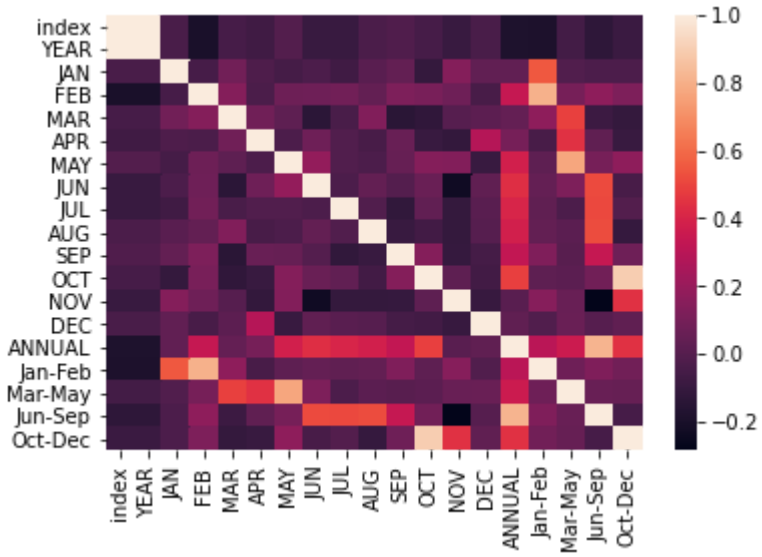


In [16]:

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

Out[16]:

```
<AxesSubplot:>
```



In [ ]: