

# 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")[1704:1817]
d
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
1704	1704	JAMMU & KASHMIR	1903	96.2	21.5	238.6	58.7	57.3	18.9	332.5	218.6	176.9
1705	1705	JAMMU & KASHMIR	1904	110.6	17.3	145.2	64.5	67.8	25.9	182.3	132.2	62.5
1706	1706	JAMMU & KASHMIR	1905	146.7	76.3	161.4	71.7	65.2	43.3	145.2	111.5	239.7
1707	1707	JAMMU & KASHMIR	1906	81.0	160.4	167.2	49.3	39.4	52.2	107.0	257.4	237.0
1708	1708	JAMMU & KASHMIR	1907	99.6	195.8	132.2	151.0	57.0	85.9	85.1	220.3	24.9
...	...	...	...	...	...	...	...	...	...	...	...	...
1812	1812	JAMMU & KASHMIR	2011	43.4	211.6	97.8	89.0	32.4	72.5	81.6	131.2	72.0
1813	1813	JAMMU & KASHMIR	2012	150.9	95.8	45.2	86.6	48.9	32.6	118.8	264.9	106.7
1814	1814	JAMMU & KASHMIR	2013	52.2	136.4	41.9	47.4	47.4	80.5	125.1	219.1	41.5
1815	1815	JAMMU & KASHMIR	2014	75.8	64.0	153.1	76.1	52.7	25.3	100.5	134.6	362.8
1816	1816	JAMMU & KASHMIR	2015	27.9	187.2	341.4	173.3	64.6	121.4	233.2	129.2	130.5

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
1704	1704	JAMMU & KASHMIR	1903	96.2	21.5	238.6	58.7	57.3	18.9	332.5	218.6	176.9
1705	1705	JAMMU & KASHMIR	1904	110.6	17.3	145.2	64.5	67.8	25.9	182.3	132.2	62.5
1706	1706	JAMMU & KASHMIR	1905	146.7	76.3	161.4	71.7	65.2	43.3	145.2	111.5	239.1
1707	1707	JAMMU & KASHMIR	1906	81.0	160.4	167.2	49.3	39.4	52.2	107.0	257.4	237.0
1708	1708	JAMMU & KASHMIR	1907	99.6	195.8	132.2	151.0	57.0	85.9	85.1	220.3	24.9
...	...	...	...	...	...	...	...	...	...	...	...	...
1812	1812	JAMMU & KASHMIR	2011	43.4	211.6	97.8	89.0	32.4	72.5	81.6	131.2	72.0
1813	1813	JAMMU & KASHMIR	2012	150.9	95.8	45.2	86.6	48.9	32.6	118.8	264.9	106.1
1814	1814	JAMMU & KASHMIR	2013	52.2	136.4	41.9	47.4	47.4	80.5	125.1	219.1	41.2
1815	1815	JAMMU & KASHMIR	2014	75.8	64.0	153.1	76.1	52.7	25.3	100.5	134.6	362.8
1816	1816	JAMMU & KASHMIR	2015	27.9	187.2	341.4	173.3	64.6	121.4	233.2	129.2	130.2

112 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-Feb',  
      'Mar-May', 'Jun-Sep', 'Oct-Dec'],  
      dtype='object')
```

In [5]:

```
d.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113 entries, 1704 to 1816
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             112 non-null    float64
10  AUG             113 non-null    float64
11  SEP             113 non-null    float64
12  OCT             113 non-null    float64
13  NOV             112 non-null    float64
14  DEC             112 non-null    float64
15  ANNUAL          112 non-null    float64
16  Jan-Feb         113 non-null    float64
17  Mar-May         113 non-null    float64
18  Jun-Sep         112 non-null    float64
19  Oct-Dec         112 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 17.8+ KB
```

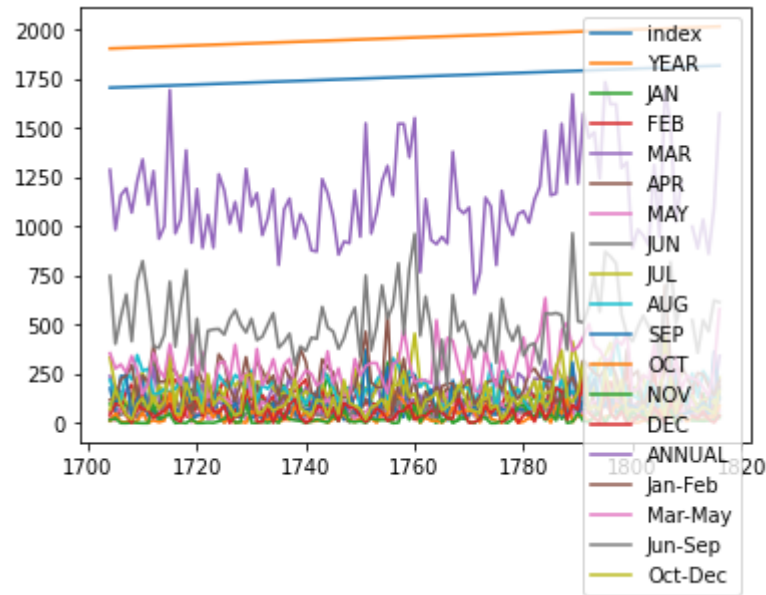
# Line Chart

In [6]:

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

Out[6]:

<AxesSubplot:>



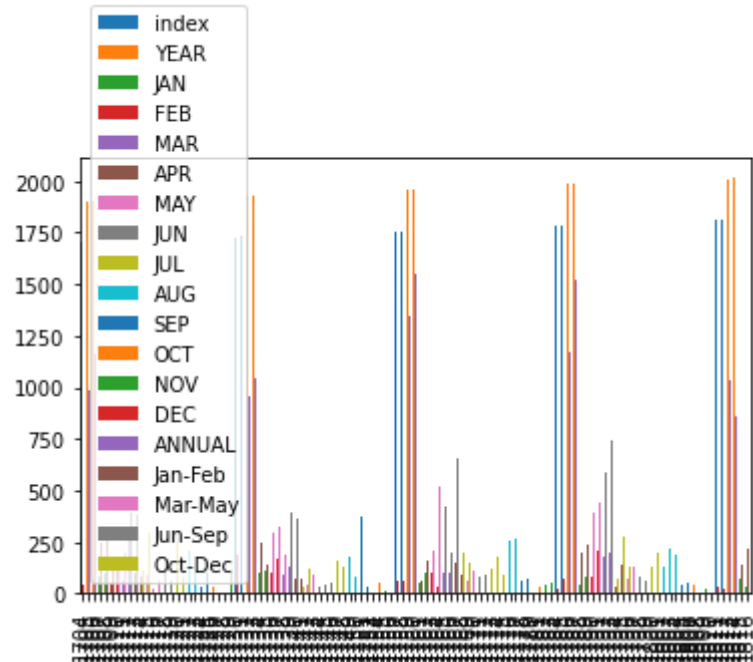
# Bar Chart

In [7]:

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

Out[7]:

<AxesSubplot:>



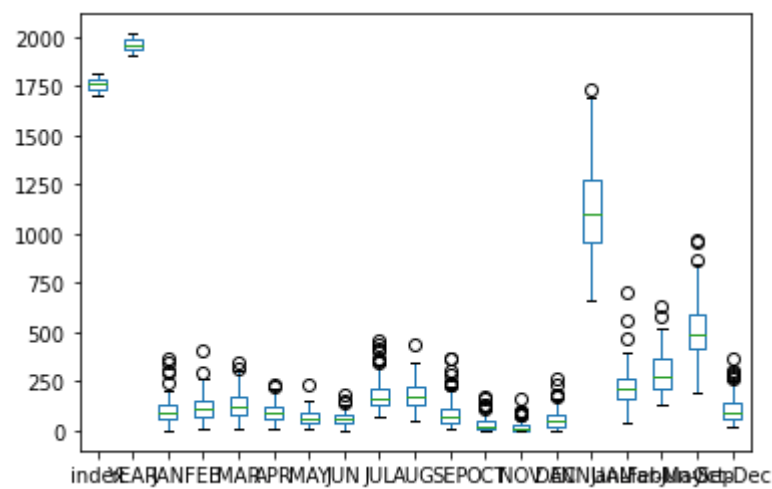
# Box Chart

In [8]:

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

Out[8]:

<AxesSubplot:>



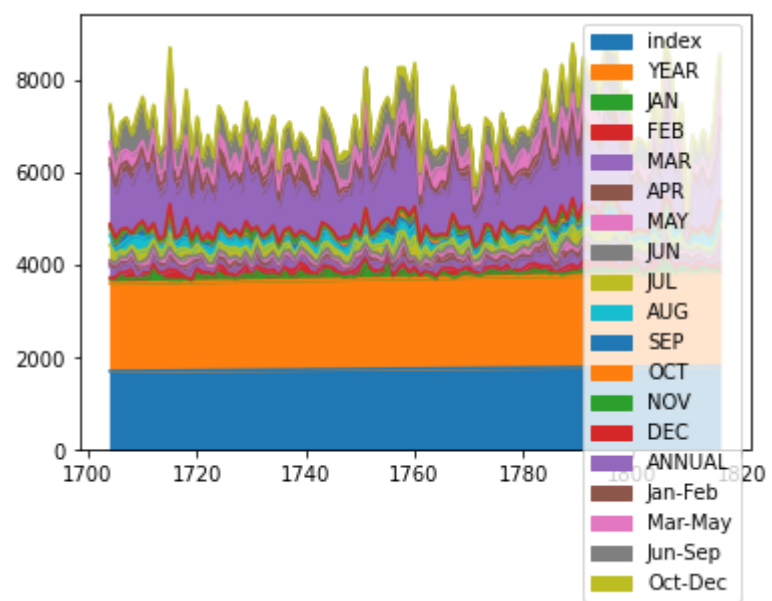
# Area Chart

In [9]:

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

Out[9]:

<AxesSubplot:>



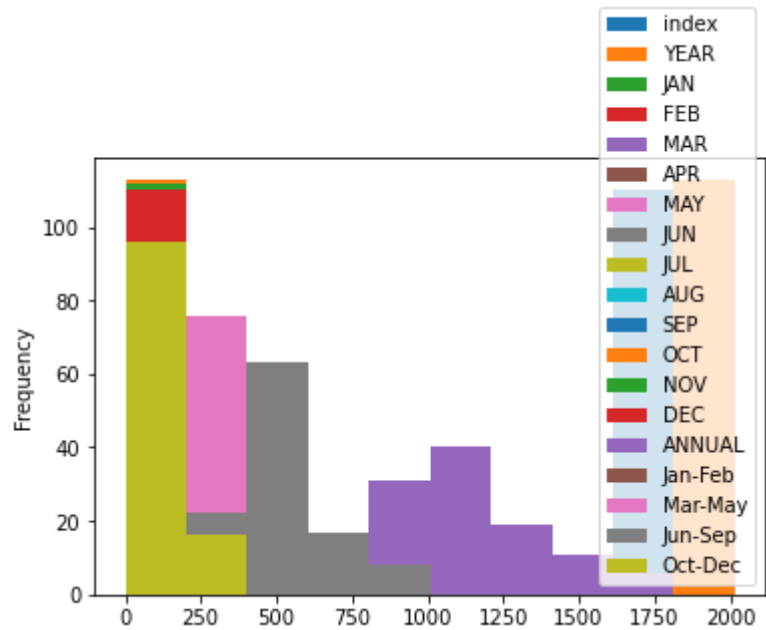
# Histogram

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[52]:

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



In [13]:

```
d.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	1760.00000	1959.00000	103.19115	116.794690	132.279646	93.30177	67.108850
std	32.76431	32.76431	66.10698	66.629372	72.006307	48.66987	37.594983
min	1704.00000	1903.00000	0.000000	9.900000	9.900000	5.700000	4.700000
25%	1732.00000	1931.00000	59.900000	66.400000	85.200000	59.100000	39.900000
50%	1760.00000	1959.00000	91.900000	106.600000	116.800000	86.600000	61.100000
75%	1788.00000	1987.00000	131.700000	150.100000	176.100000	119.300000	91.600000
max	1816.00000	2015.00000	367.800000	403.500000	341.400000	233.200000	234.400000

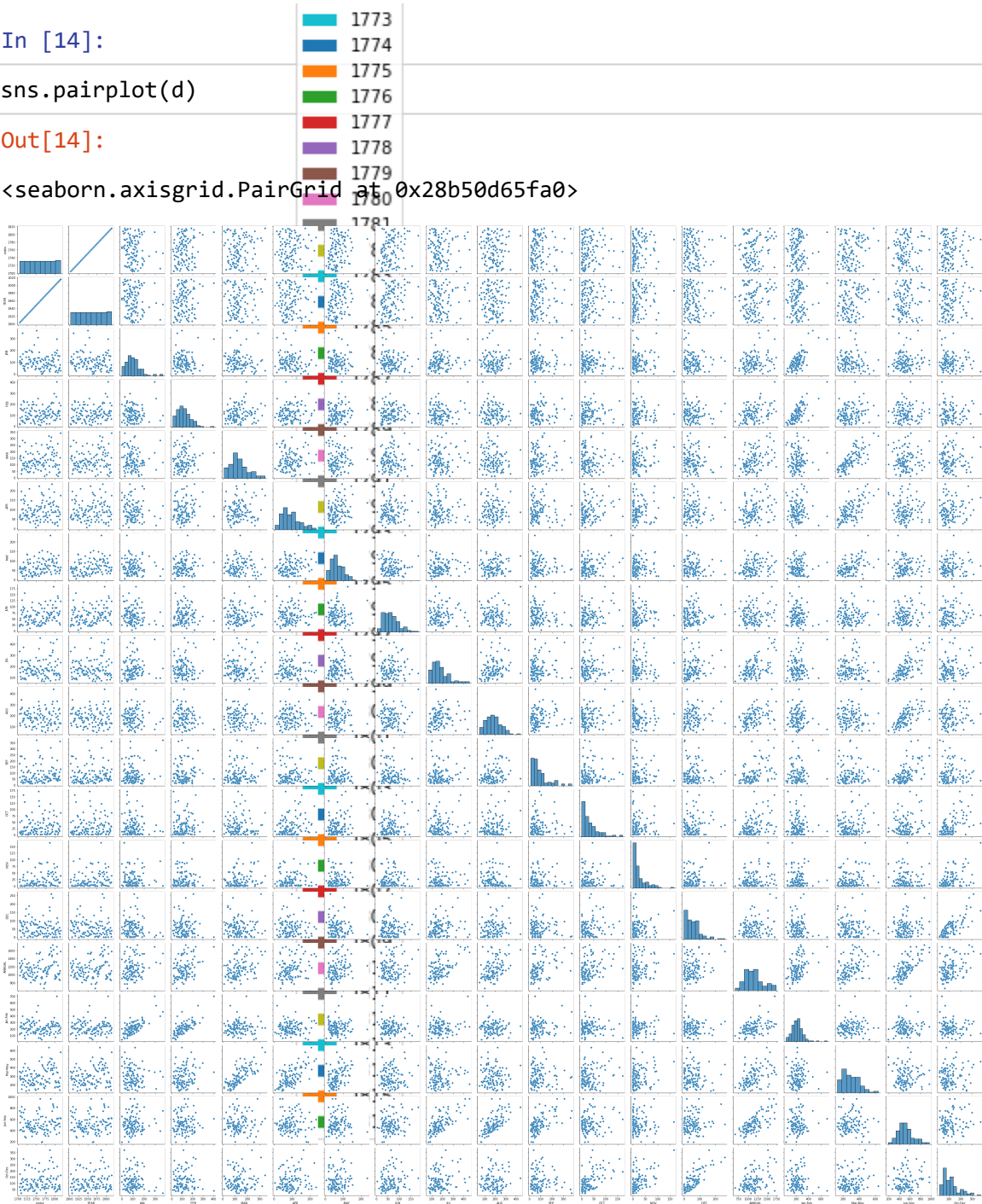
# EDA AND VISUALIZATION

In [14]:

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

Out[14]:

<seaborn.axisgrid.PairGrid at 0x28b50d65fa0>

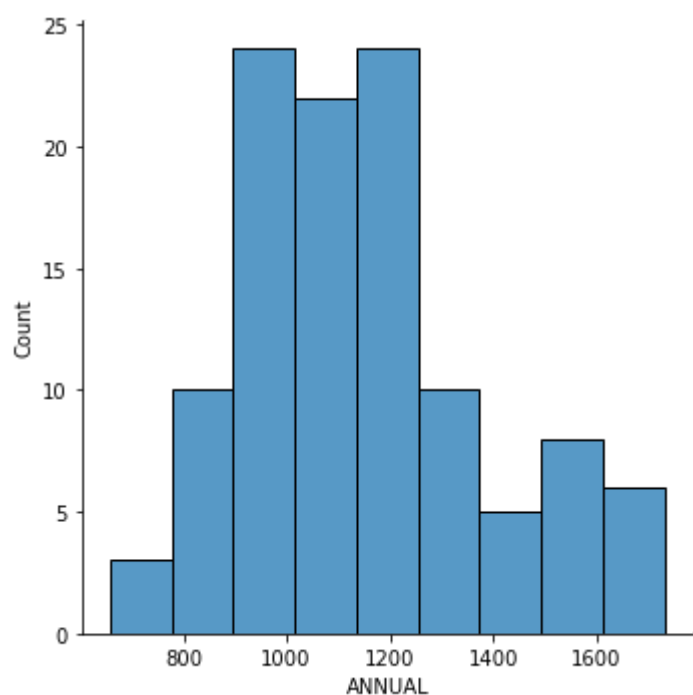


In [15]:

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

Out[15]:

```
<seaborn.axisgrid.FacetGrid at 0x28b5badabb0>
```

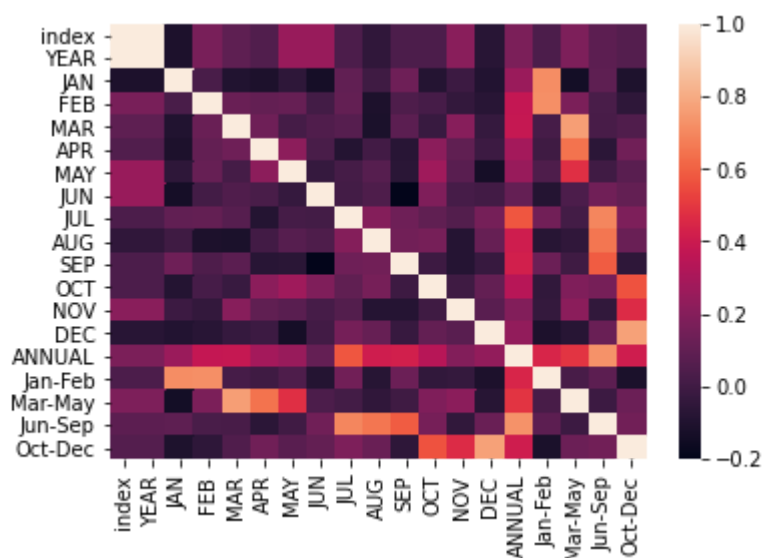


In [16]:

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

Out[16]:

```
<AxesSubplot:>
```



In [ ]:

