Import Libraries

In [1]:

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

In [2]:

```python
d=pd.read_csv(r"C:\Users\user\Downloads\FP2_RainFall\rain.csv")[4003:4116]
d
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEF
4003	4003	LAKSHADWEEP	1902	99.3	9.6	32.6	40.4	179.1	374.2	413.3	170.0	214.3
4004	4004	LAKSHADWEEP	1903	63.5	95.0	0.0	29.5	144.1	212.4	261.8	202.0	292.1
4005	4005	LAKSHADWEEP	1904	0.0	0.0	13.5	13.2	143.3	261.3	256.0	38.9	219.9
4006	4006	LAKSHADWEEP	1905	62.4	0.0	0.0	0.0	166.7	400.7	68.7	377.5	107.5
4007	4007	LAKSHADWEEP	1906	17.8	0.0	24.4	33.8	213.0	465.0	348.6	260.5	25.9
...	...	...	...	...	...	...	...	...	...	...	...	..
4111	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2
4112	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8
4113	4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.0
4114	4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.2
4115	4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	296.6	257.5	146.4	160.4

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	SE
4003	4003	LAKSHADWEEP	1902	99.3	9.6	32.6	40.4	179.1	374.2	413.3	170.0	214.
4005	4005	LAKSHADWEEP	1904	0.0	0.0	13.5	13.2	143.3	261.3	256.0	38.9	219.
4006	4006	LAKSHADWEEP	1905	62.4	0.0	0.0	0.0	166.7	400.7	68.7	377.5	107.
4007	4007	LAKSHADWEEP	1906	17.8	0.0	24.4	33.8	213.0	465.0	348.6	260.5	25.
4008	4008	LAKSHADWEEP	1907	60.6	49.3	0.0	123.5	77.0	241.1	199.5	165.6	25.
...	...	...	...	...	...	...	...	...	...	...	...	.
4111	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.
4112	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.
4113	4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.
4114	4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.
4115	4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	296.6	257.5	146.4	160.

102 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, 4003 to 4115
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          111 non-null    float64
 4   FEB          112 non-null    float64
 5   MAR          111 non-null    float64
 6   APR          111 non-null    float64
 7   MAY          111 non-null    float64
 8   JUN          111 non-null    float64
 9   JUL          110 non-null    float64
 10  AUG          111 non-null    float64
 11  SEP          110 non-null    float64
 12  OCT          110 non-null    float64
 13  NOV          107 non-null    float64
 14  DEC          109 non-null    float64
 15  ANNUAL       102 non-null    float64
 16  Jan-Feb      110 non-null    float64
 17  Mar-May      109 non-null    float64
 18  Jun-Sep      109 non-null    float64
 19  Oct-Dec      107 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 17.8+ KB
```
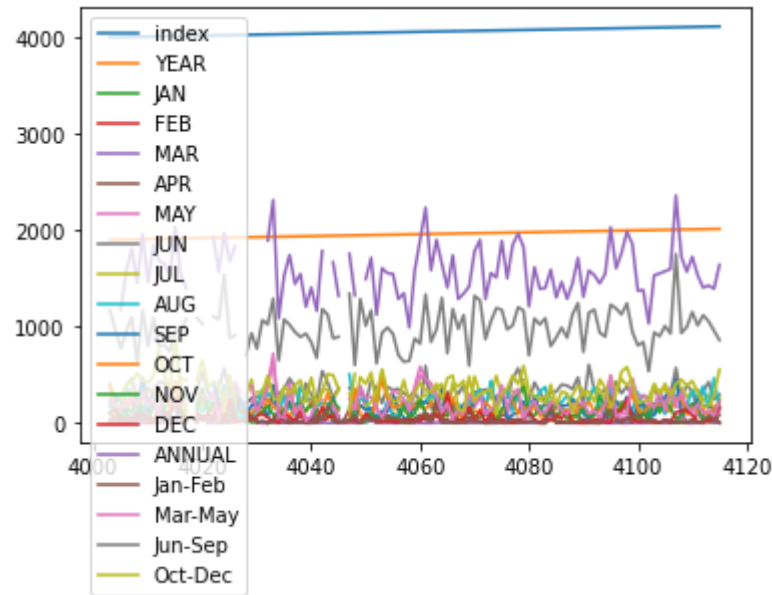
Line Chart
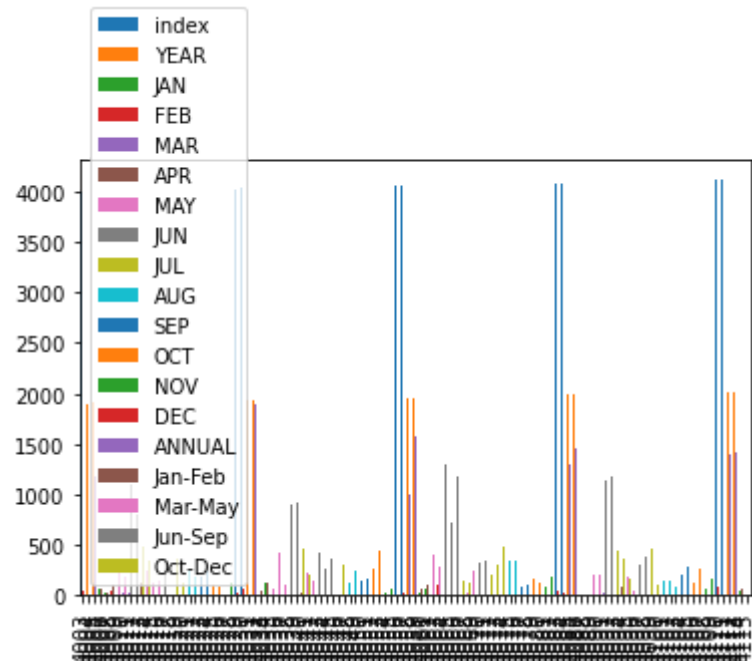
In [6]:

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

Out[6]:

`<AxesSubplot:>`
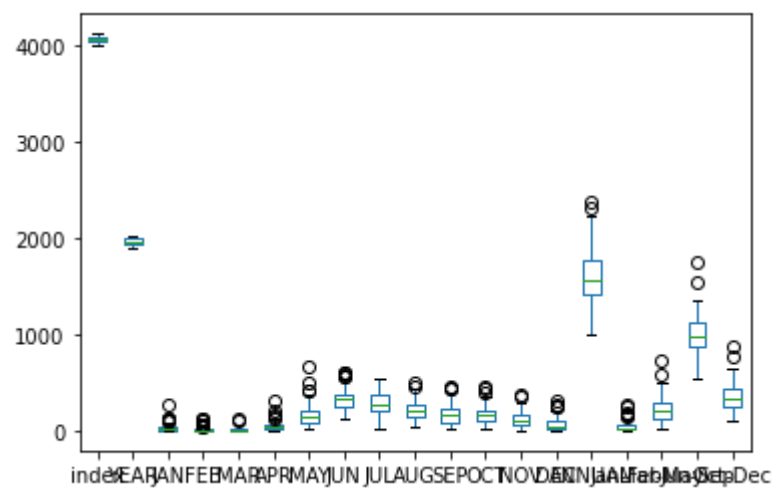

Bar Chart

In [7]:

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

Out[7]:

`<AxesSubplot:>`

Box Chart

In [8]:
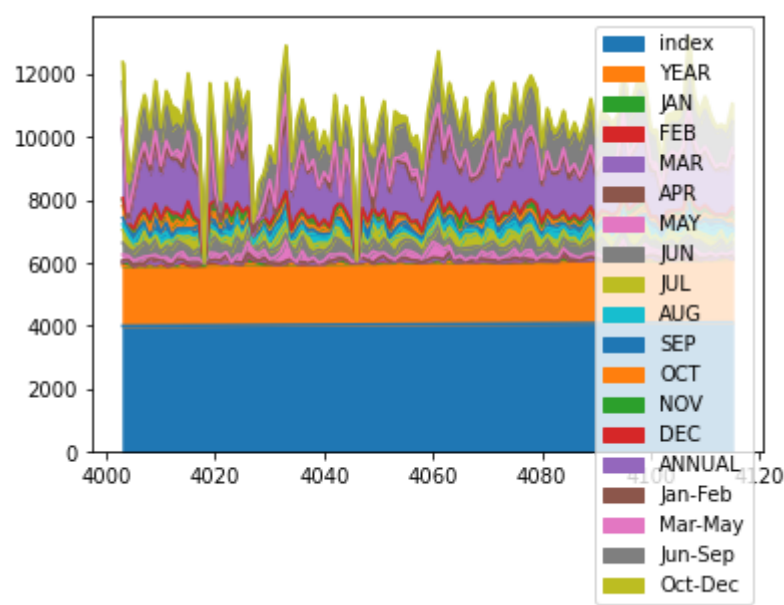
```
d.plot.box()
```

Out[8]:

```
<AxesSubplot:>
```


Area Chart

In [9]:
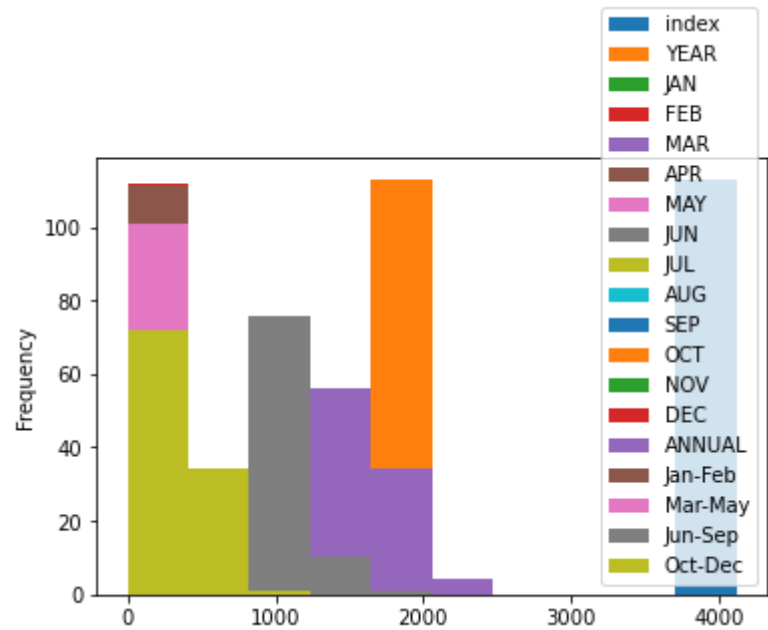
```
d.plot.area()
```

Out[9]:

```
<AxesSubplot:>
```


Histogram

In [10]:

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

Out[10]:

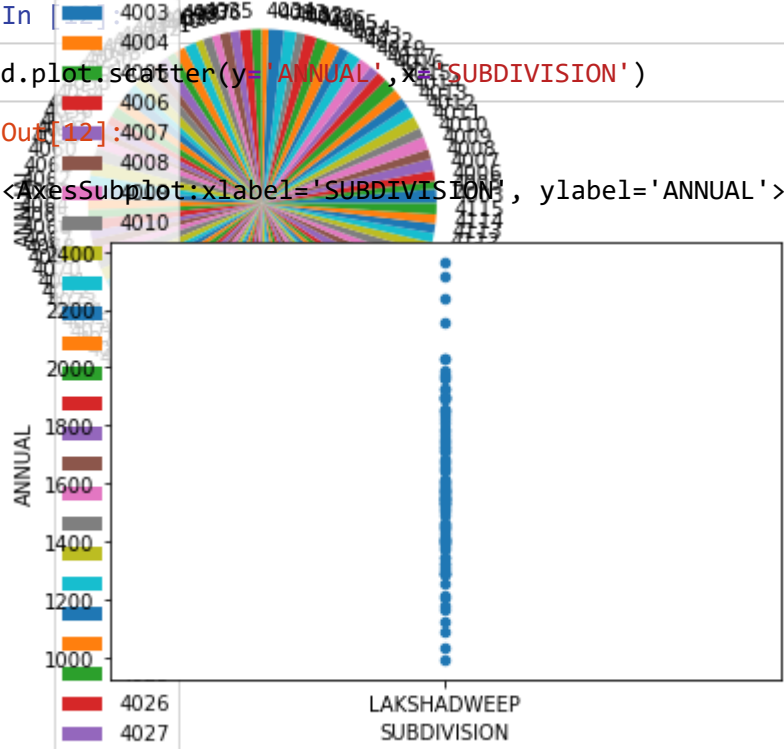`<AxesSubplot:ylabel='Frequency'>`

Pie Chart

In [11]:

```python
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	113.000000	111.000000	112.000000	111.000000	111.000000	111.000000
mean	4059.00000	1958.858407	27.538739	15.204464	13.445946	43.193694	165.034234
std	32.76431	32.976958	38.315053	23.950077	19.138361	46.565950	109.090756
min	4003.00000	1902.000000	0.000000	0.000000	0.000000	0.000000	13.500000
25%	4031.00000	1931.000000	4.200000	0.400000	0.200000	14.050000	88.100000
50%	4059.00000	1959.000000	12.400000	3.950000	5.200000	32.300000	148.300000
75%	4087.00000	1987.000000	38.250000	18.275000	22.150000	58.900000	210.450000
max	4115.00000	2015.000000	262.800000	114.900000	120.700000	315.400000	660.800000

EDA AND VISUALIZATION

In [14]:

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

Out[14]:

```
<seaborn.axisgrid.PairGrid at 0x1f3f2320190>
```
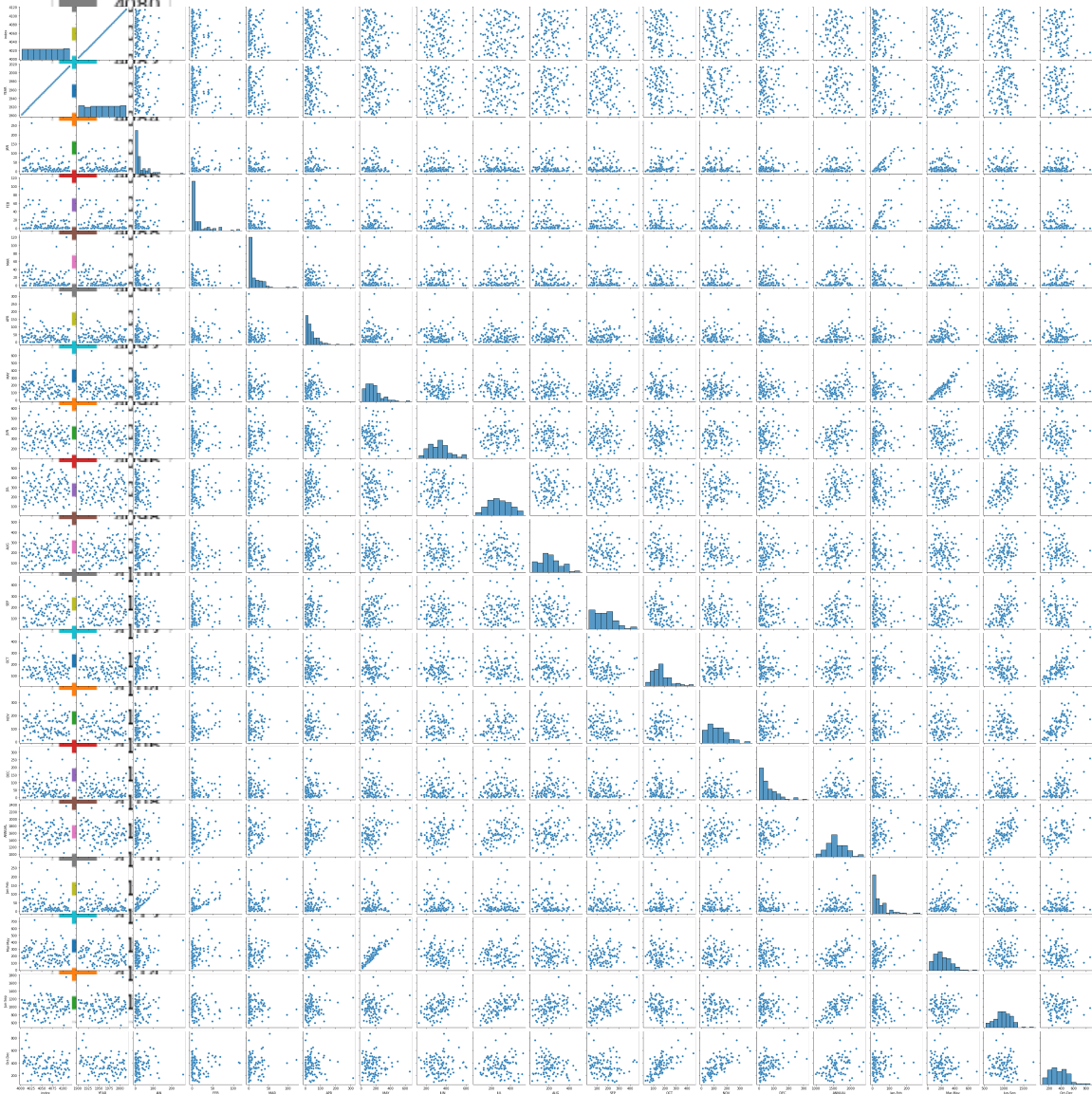
In [15]:

```python
sns.displot(d['ANNUAL'])
```
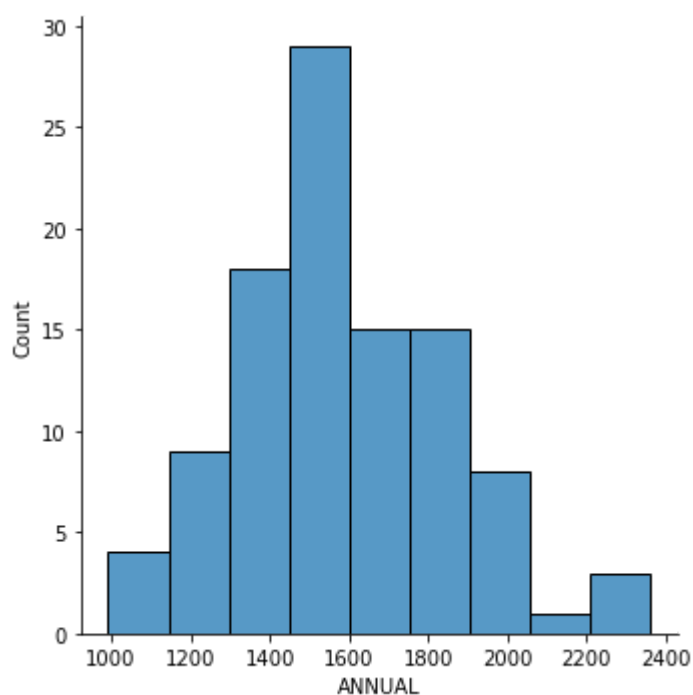
Out[15]:
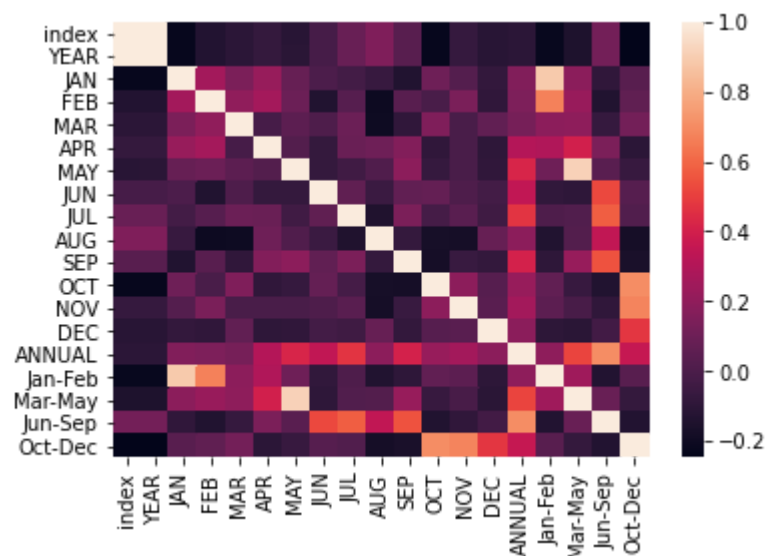
```
<seaborn.axisgrid.FacetGrid at 0x1f3fe779970>
```


In [16]:

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

Out[16]:

```
<AxesSubplot:>
```


In [ ]: