

# Import Libraries

In [1]:

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

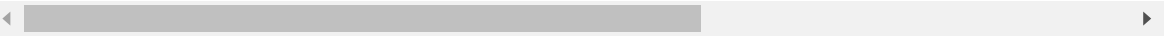
In [2]:

```
d=pd.read_csv(r"C:\Users\user\Downloads\FP2_RainFall\rain.csv")[3888:4002]
d
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SE
3888	3888	KERALA	1902	6.7	2.6	57.3	83.9	134.5	390.9	1205.0	315.8	491.
3889	3889	KERALA	1903	3.2	18.6	3.1	83.6	249.7	558.6	1022.5	420.2	341.
3890	3890	KERALA	1904	23.7	3.0	32.2	71.5	235.7	1098.2	725.5	351.8	222.
3891	3891	KERALA	1905	1.2	22.3	9.4	105.9	263.3	850.2	520.5	293.6	217.
3892	3892	KERALA	1906	26.7	7.4	9.9	59.4	160.8	414.9	954.2	442.8	131.
...	...	...	...	...	...	...	...	...	...	...	...	.
3997	3997	KERALA	2011	20.5	45.7	24.1	165.2	124.2	788.5	536.8	492.7	391.
3998	3998	KERALA	2012	7.4	11.0	21.0	171.1	95.3	430.3	362.6	501.6	241.
3999	3999	KERALA	2013	3.9	40.1	49.9	49.3	119.3	1042.7	830.2	369.7	318.
4000	4000	KERALA	2014	4.6	10.3	17.9	95.7	251.0	454.4	677.8	733.9	298.
4001	4001	KERALA	2015	3.1	5.8	50.1	214.1	201.8	563.6	406.0	252.2	292.

114 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
3888	3888	KERALA	1902	6.7	2.6	57.3	83.9	134.5	390.9	1205.0	315.8	491.
3889	3889	KERALA	1903	3.2	18.6	3.1	83.6	249.7	558.6	1022.5	420.2	341.
3890	3890	KERALA	1904	23.7	3.0	32.2	71.5	235.7	1098.2	725.5	351.8	222.
3891	3891	KERALA	1905	1.2	22.3	9.4	105.9	263.3	850.2	520.5	293.6	217.
3892	3892	KERALA	1906	26.7	7.4	9.9	59.4	160.8	414.9	954.2	442.8	131.
...	...	...	...	...	...	...	...	...	...	...	...	.
3997	3997	KERALA	2011	20.5	45.7	24.1	165.2	124.2	788.5	536.8	492.7	391.
3998	3998	KERALA	2012	7.4	11.0	21.0	171.1	95.3	430.3	362.6	501.6	241.
3999	3999	KERALA	2013	3.9	40.1	49.9	49.3	119.3	1042.7	830.2	369.7	318.
4000	4000	KERALA	2014	4.6	10.3	17.9	95.7	251.0	454.4	677.8	733.9	298.
4001	4001	KERALA	2015	3.1	5.8	50.1	214.1	201.8	563.6	406.0	252.2	292.

114 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: 114 entries, 3888 to 4001
Data columns (total 20 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   index           114 non-null    int64
 1   SUBDIVISION     114 non-null    object
 2   YEAR            114 non-null    int64
 3   JAN             114 non-null    float64
 4   FEB             114 non-null    float64
 5   MAR             114 non-null    float64
 6   APR             114 non-null    float64
 7   MAY             114 non-null    float64
 8   JUN             114 non-null    float64
 9   JUL             114 non-null    float64
10  AUG             114 non-null    float64
11  SEP             114 non-null    float64
12  OCT             114 non-null    float64
13  NOV             114 non-null    float64
14  DEC             114 non-null    float64
15  ANNUAL          114 non-null    float64
16  Jan-Feb         114 non-null    float64
17  Mar-May         114 non-null    float64
18  Jun-Sep         114 non-null    float64
19  Oct-Dec         114 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 17.9+ KB
```

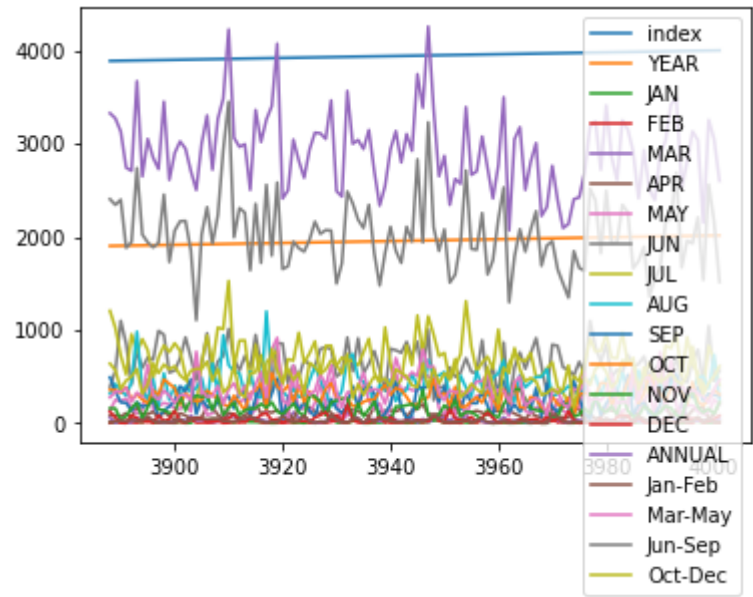
# Line Chart

In [6]:

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

Out[6]:

<AxesSubplot:>



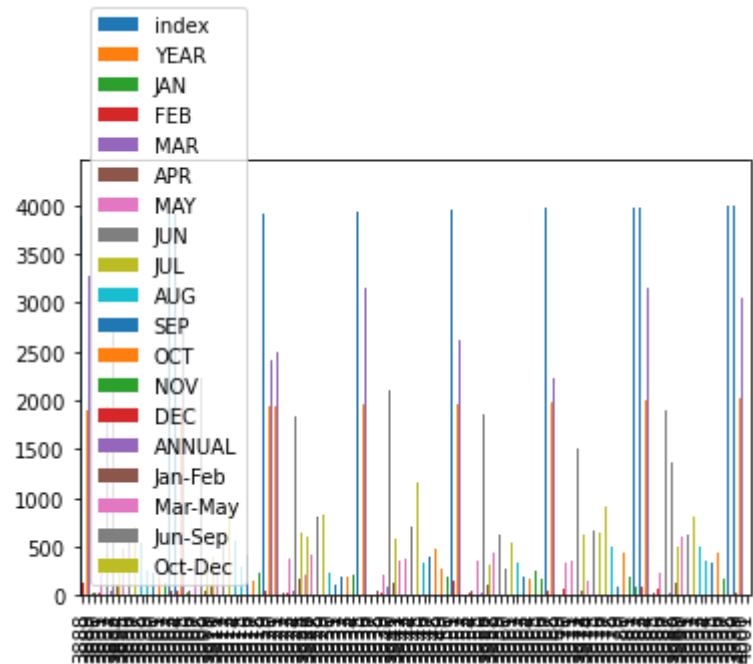
# Bar Chart

In [7]:

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

Out[7]:

<AxesSubplot:>

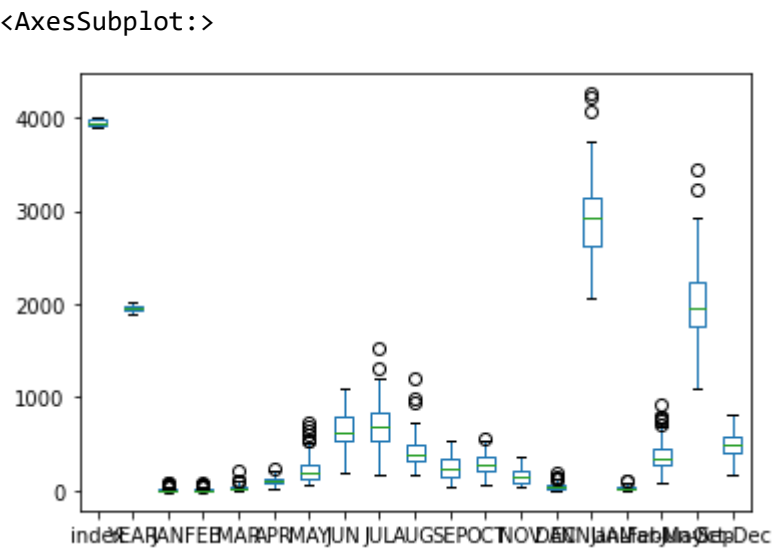


# Box Chart

In [8]:

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

Out[8]:

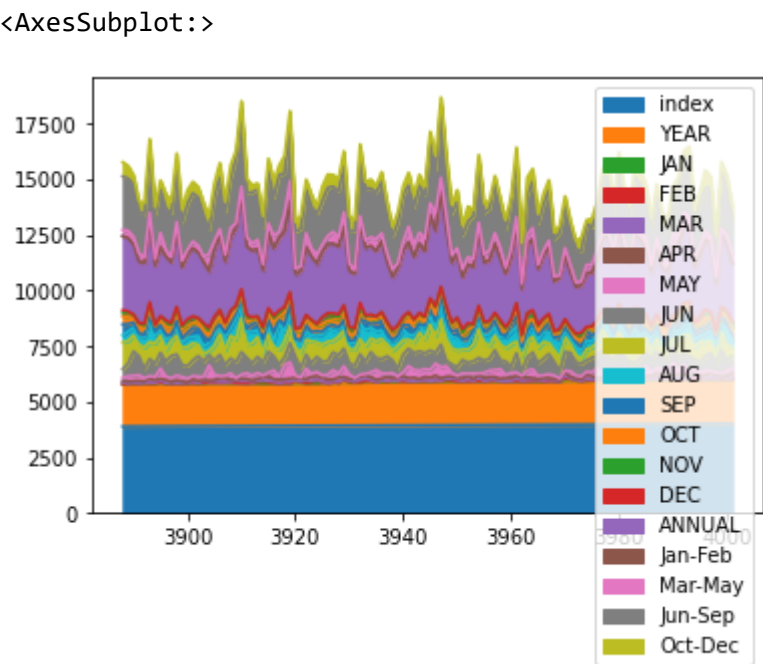


# Area Chart

In [9]:

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

Out[9]:



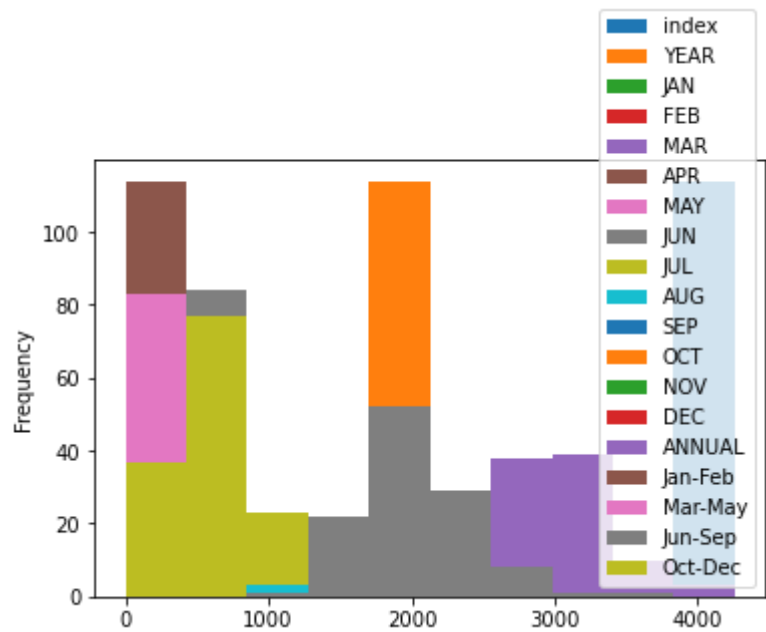
# Histogram

In [10]:

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

Out[10]:

<AxesSubplot:ylabel='Frequency'>



# Pie Chart

In [11]:

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

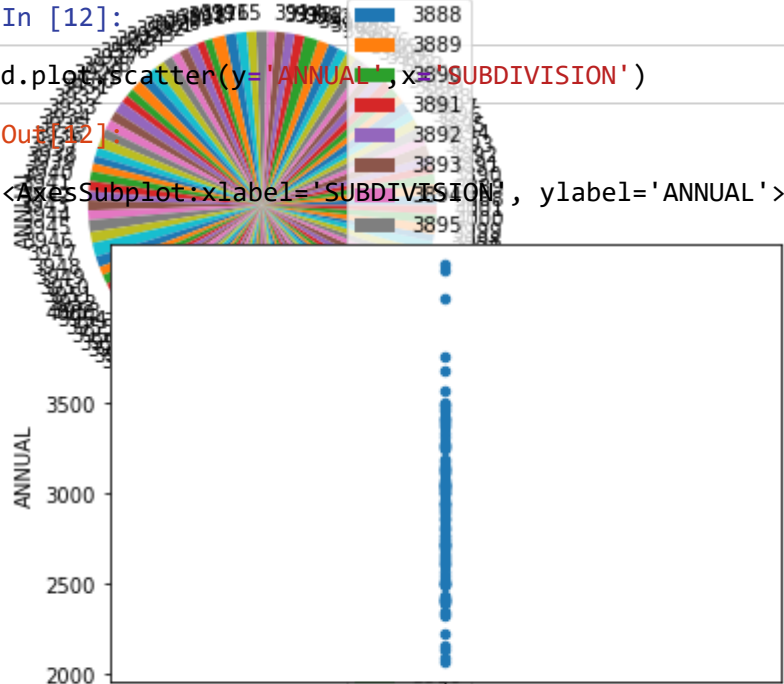
Out[11]:

<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	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000
mean	3944.500000	1958.500000	12.102632	15.240351	36.685088	110.140351	230.365789
std	33.052988	33.052988	15.529917	16.042564	30.426433	44.627534	149.840056
min	3888.000000	1902.000000	0.000000	0.000000	0.100000	13.100000	53.400000
25%	3916.250000	1930.250000	2.175000	4.700000	18.100000	74.350000	124.275000
50%	3944.500000	1958.500000	5.800000	8.350000	28.300000	109.100000	188.000000
75%	3972.750000	1986.750000	16.800000	21.400000	49.825000	135.050000	283.175000
max	4001.000000	2015.000000	83.500000	79.000000	217.200000	238.000000	738.800000

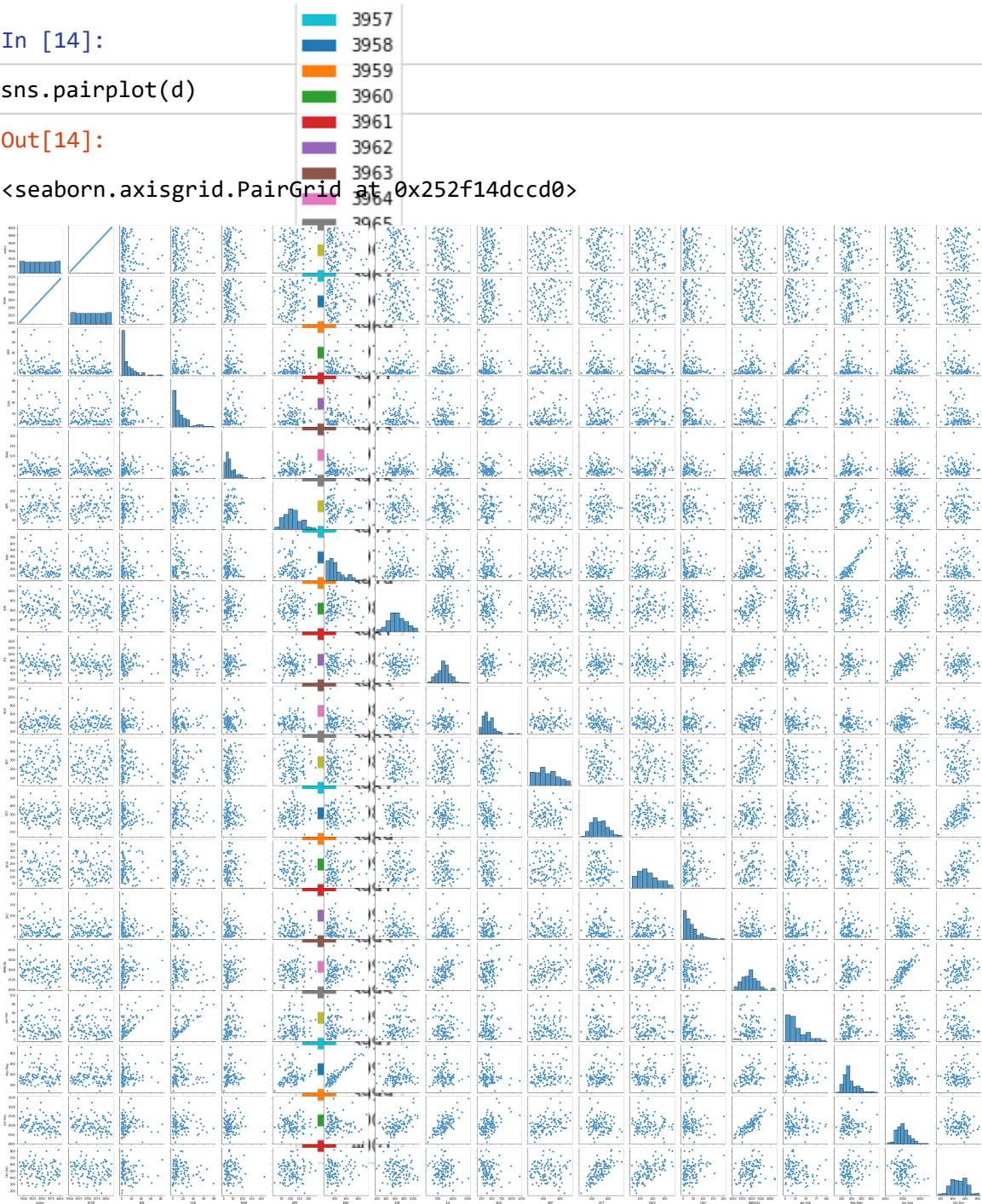
## EDA AND VISUALIZATION

In [14]:

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

Out[14]:

<seaborn.axisgrid.PairGrid at 0x252f14dcd0>

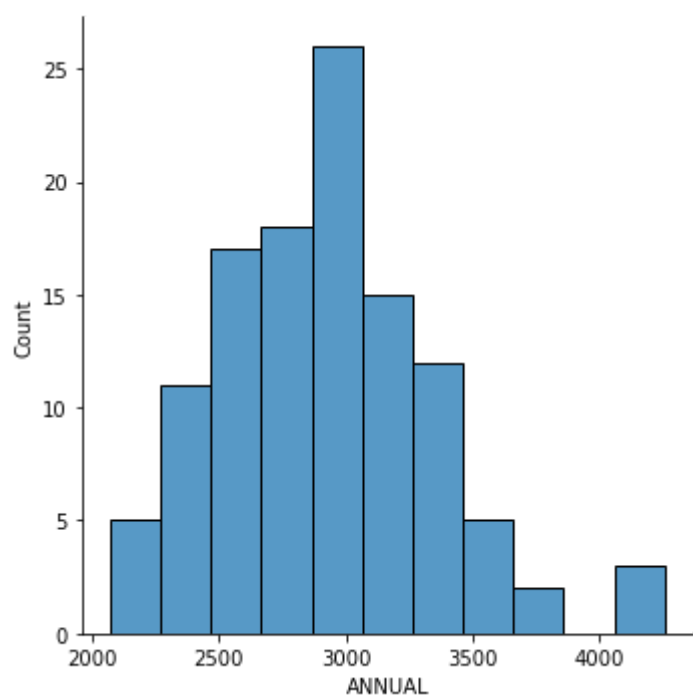


In [15]:

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

Out[15]:

```
<seaborn.axisgrid.FacetGrid at 0x252fd93ebb0>
```

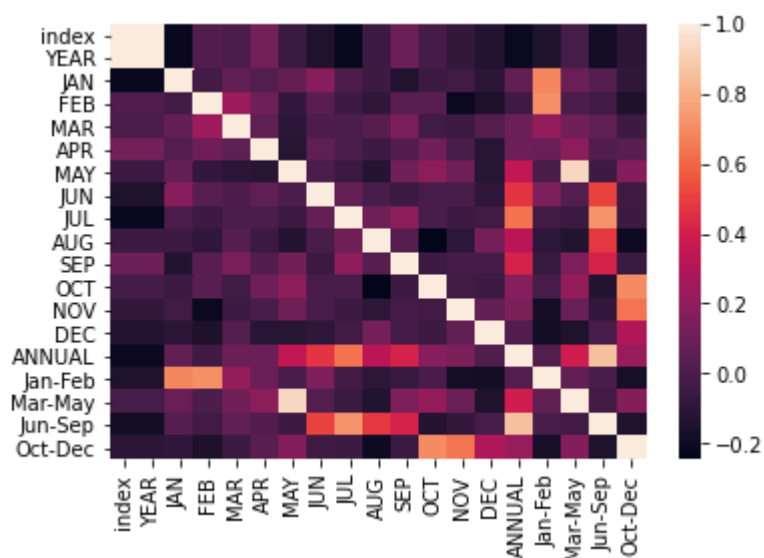


In [16]:

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

Out[16]:

```
<AxesSubplot:>
```



In [ ]:

