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
df=pd.read_csv(r"c:\Users\user\Downloads\FP2_RainFall\rainfall.csv")[554:667]
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
554	554	GANGETIC WEST BENGAL	1903	17.5	24.6	37.3	30.6	78.5	201.7	179.6	277.6	300.7
555	555	GANGETIC WEST BENGAL	1904	0.1	23.9	35.6	17.5	160.2	286.7	435.3	241.7	142.8
556	556	GANGETIC WEST BENGAL	1905	30.9	49.6	84.7	84.9	156.8	70.9	525.5	263.6	287.6
557	557	GANGETIC WEST BENGAL	1906	46.8	123.6	53.3	1.1	81.3	172.0	307.1	264.6	193.2
558	558	GANGETIC WEST BENGAL	1907	0.7	24.2	94.9	39.7	92.3	347.6	217.5	249.9	232.3
...	...	...	...	...	...	...	...	...	...	...	...	...
662	662	GANGETIC WEST BENGAL	2011	2.5	2.7	40.5	75.0	132.6	434.5	219.9	443.2	295.9
663	663	GANGETIC WEST BENGAL	2012	40.7	15.3	4.4	57.7	44.2	146.6	315.0	261.4	246.9
664	664	GANGETIC WEST BENGAL	2013	2.5	10.0	4.8	45.6	195.9	233.4	263.2	401.4	254.0
665	665	GANGETIC WEST BENGAL	2014	0.9	42.2	19.9	1.9	124.4	193.6	298.7	292.6	229.5
666	666	GANGETIC WEST BENGAL	2015	12.9	5.5	19.3	88.7	57.6	247.2	633.1	260.6	164.0

113 rows × 20 columns

Data Cleaning and Preprocessing

In [3]:

```python
df.dropna()
```

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
554	554	GANGETIC WEST BENGAL	1903	17.5	24.6	37.3	30.6	78.5	201.7	179.6	277.6	300.7
555	555	GANGETIC WEST BENGAL	1904	0.1	23.9	35.6	17.5	160.2	286.7	435.3	241.7	142.8
556	556	GANGETIC WEST BENGAL	1905	30.9	49.6	84.7	84.9	156.8	70.9	525.5	263.6	287.6
557	557	GANGETIC WEST BENGAL	1906	46.8	123.6	53.3	1.1	81.3	172.0	307.1	264.6	193.2
558	558	GANGETIC WEST BENGAL	1907	0.7	24.2	94.9	39.7	92.3	347.6	217.5	249.9	232.3
...	...	...	...	...	...	...	...	...	...	...	...	...
662	662	GANGETIC WEST BENGAL	2011	2.5	2.7	40.5	75.0	132.6	434.5	219.9	443.2	295.9
663	663	GANGETIC WEST BENGAL	2012	40.7	15.3	4.4	57.7	44.2	146.6	315.0	261.4	246.9
664	664	GANGETIC WEST BENGAL	2013	2.5	10.0	4.8	45.6	195.9	233.4	263.2	401.4	254.0
665	665	GANGETIC WEST BENGAL	2014	0.9	42.2	19.9	1.9	124.4	193.6	298.7	292.6	229.5
666	666	GANGETIC WEST BENGAL	2015	12.9	5.5	19.3	88.7	57.6	247.2	633.1	260.6	164.0

113 rows × 20 columns

In [4]:

```python
df.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]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113 entries, 554 to 666
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
```
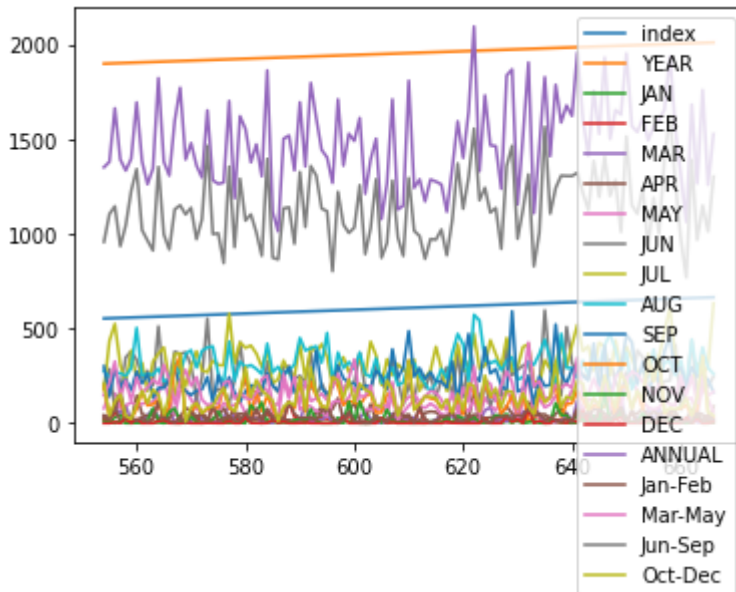
Line Chart
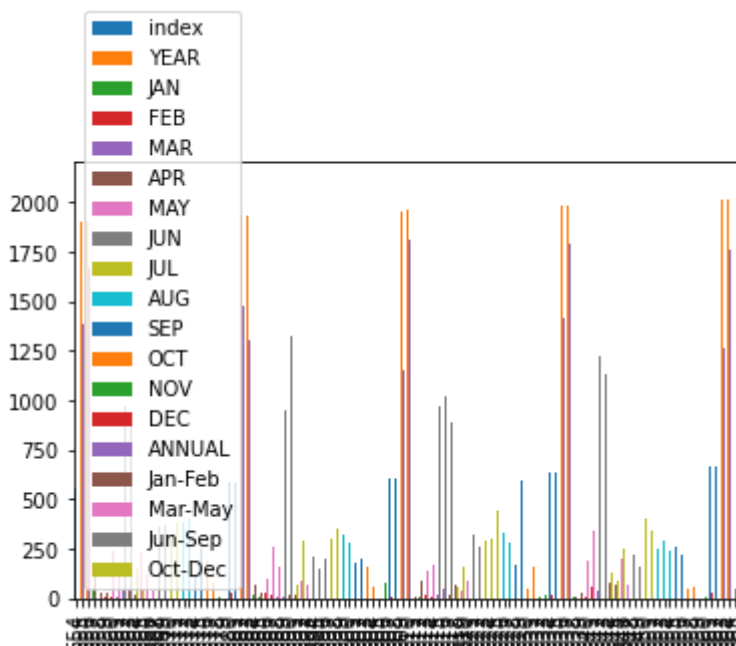
In [6]:

```
df.plot.line()
```

Out[6]:

<AxesSubplot:>

Bar chart

In [7]:

```
df.plot.bar()
```

Out[7]:

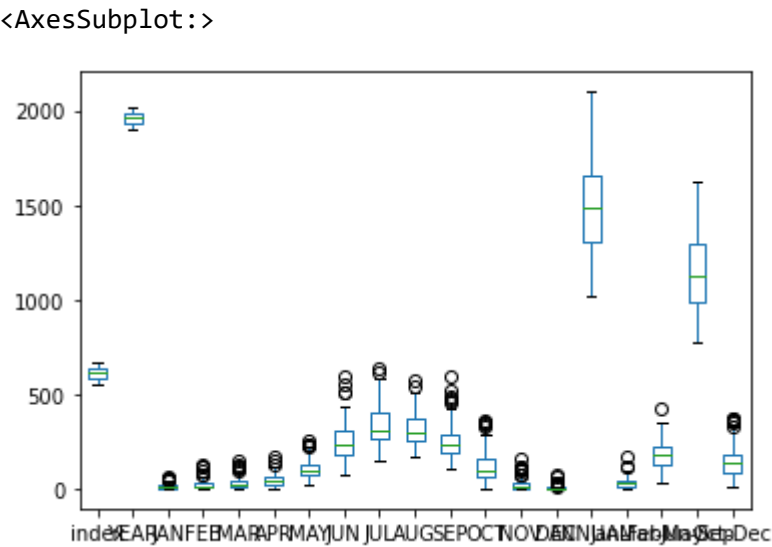<AxesSubplot:>

Box chart
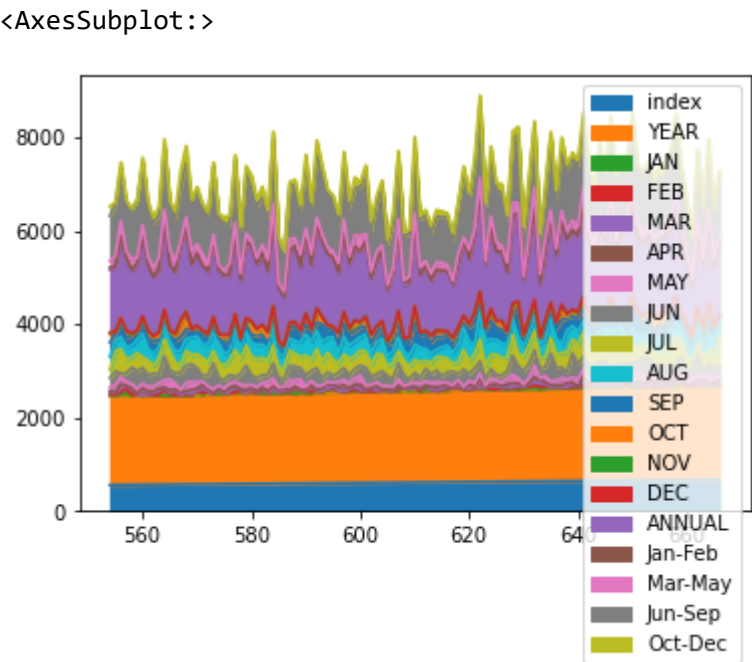
In [8]:

```
df.plot.box()
```

Out[8]:

<AxesSubplot:>

Area Chart
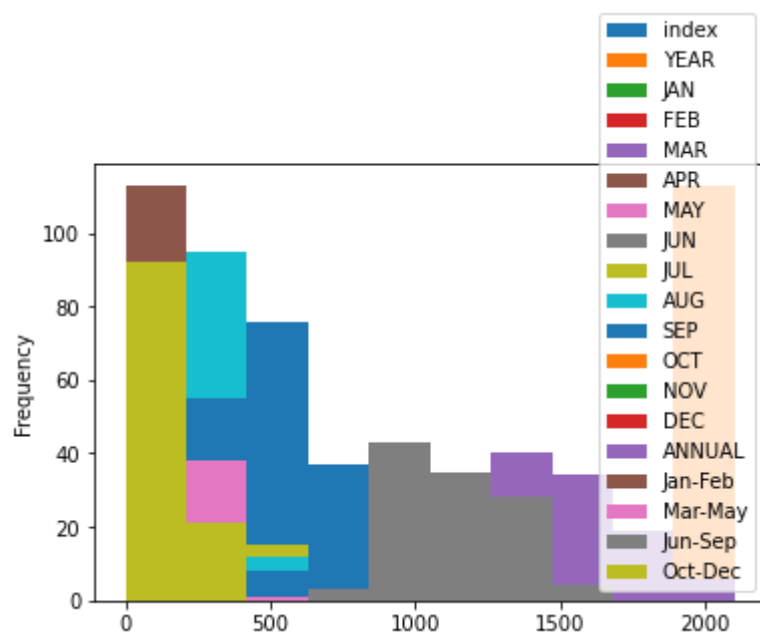
In [9]:

```
df.plot.area()
```

Out[9]:

<AxesSubplot:>

Histogram

In [10]:

```
df.plot.hist()
```

Out[10]:

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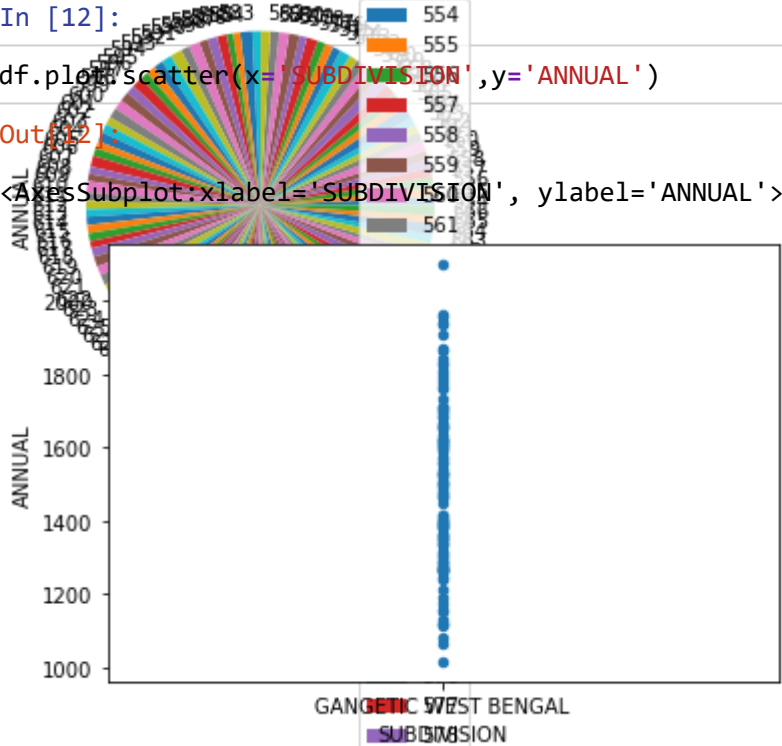

pie chart

In [11]:

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

Out[11]:

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

Scatter chart

In [12]:

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

Out[12]:

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


In [13]:

```python
df.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	610.00000	1959.00000	12.490265	22.322124	29.179646	44.193805	107.188496	24
std	32.76431	32.76431	14.642849	24.058937	30.903265	31.549460	51.182525	9
min	554.00000	1903.00000	0.000000	0.000000	0.100000	0.900000	16.400000	6
25%	582.00000	1931.00000	1.300000	5.300000	7.300000	20.200000	71.500000	18
50%	610.00000	1959.00000	6.800000	13.600000	18.900000	38.700000	98.700000	22
75%	638.00000	1987.00000	18.000000	30.300000	41.700000	58.300000	130.300000	30
max	666.00000	2015.00000	60.000000	123.600000	152.500000	174.200000	250.900000	59

EDA and Visualization
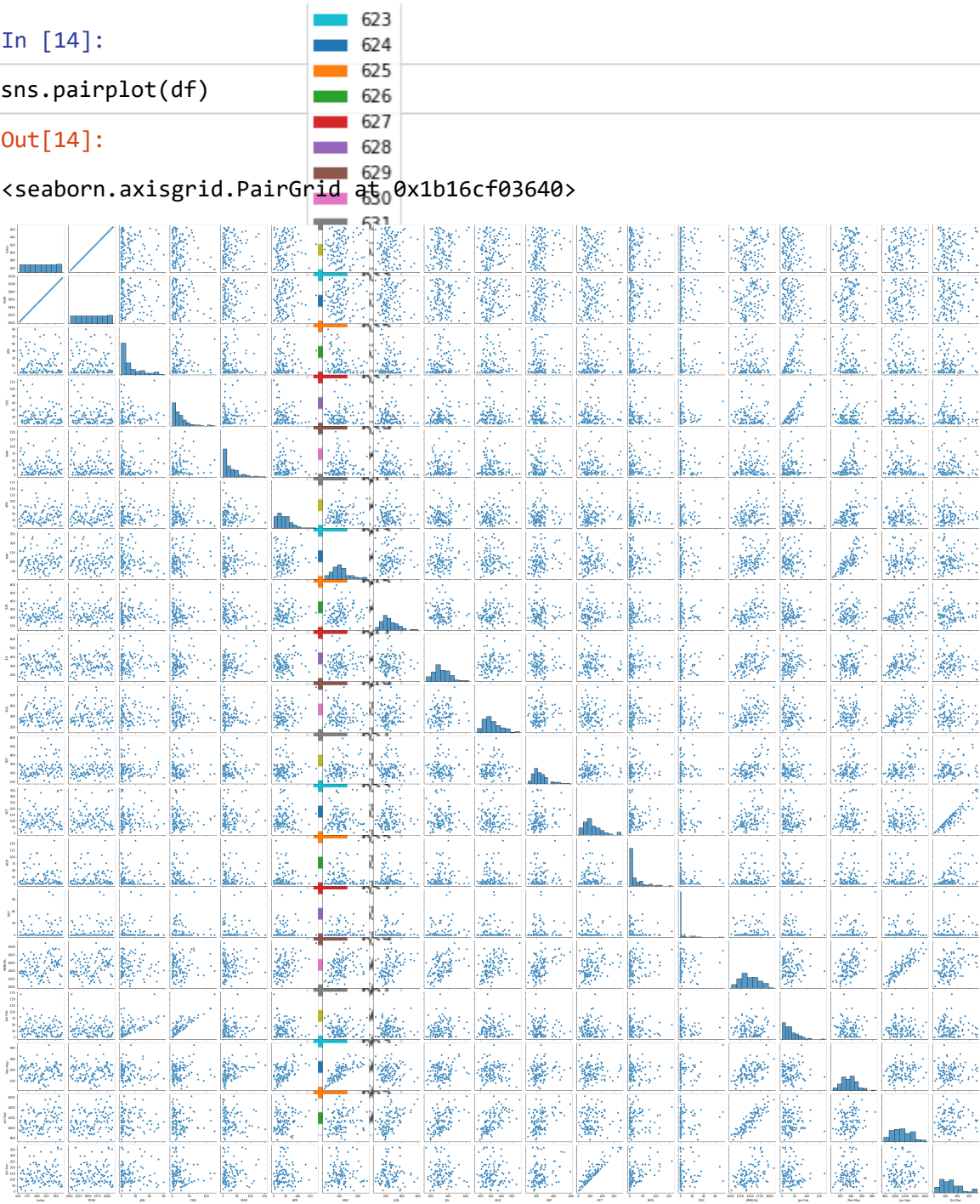
In [14]:

```
sns.pairplot(df)
```

Out[14]:

<seaborn.axisgrid.PairGrid at 0x1b16cf03640>

In [15]:

```python
sns.displot(df['ANNUAL'])
```
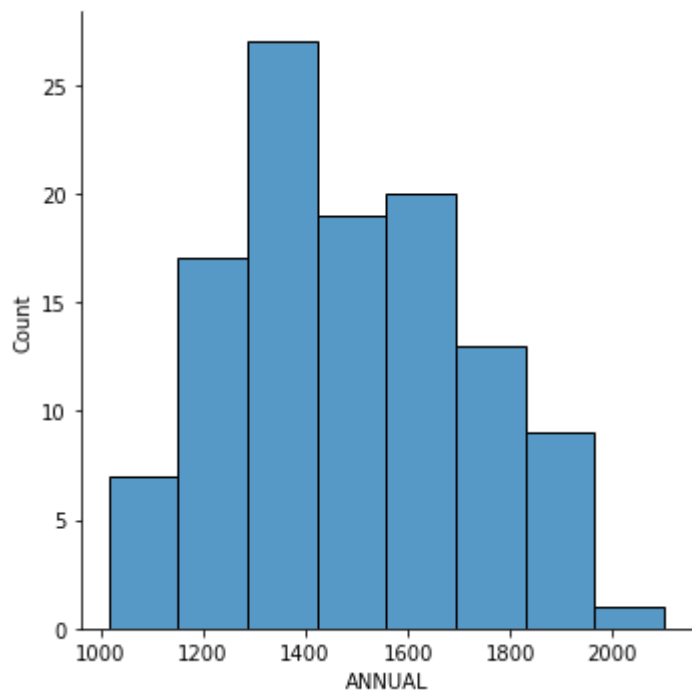
Out[15]:

```
<seaborn.axisgrid.FacetGrid at 0x1b1790325e0>
```
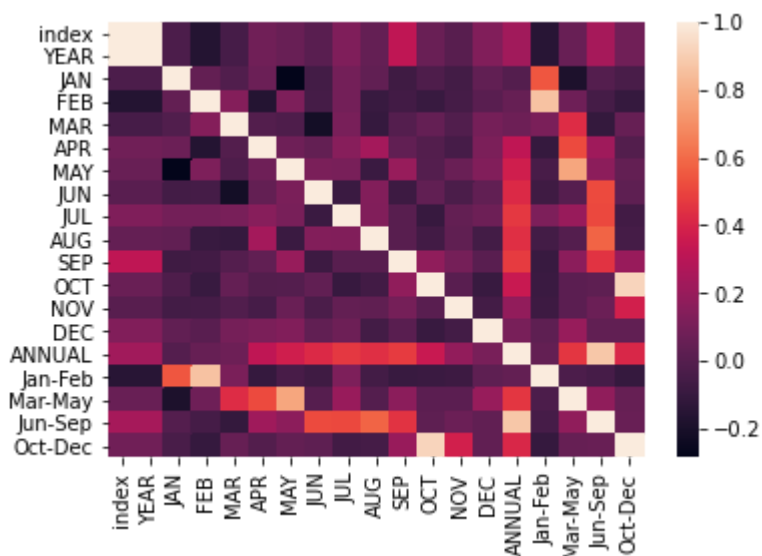

In [16]:

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

Out[16]:

```
<AxesSubplot:>
```


In [ ]: