# Import Libraries

In [1]:

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

In [2]:

```
df=pd.read_csv(r"c:\Users\user\Downloads\FP2_RainFall\rainfall.csv")[2969:3082]
df
```

Out[2]:

| | index | SUBDIVISION | YEAR | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2969** | 2969 | CHHATTISGARH | 1903 | 6.2 | 13.9 | 0.4 | 6.8 | 51.1 | 110.7 | 365.9 | 396.0 | 212.0 |
| **2970** | 2970 | CHHATTISGARH | 1904 | 0.0 | 8.6 | 32.3 | 0.2 | 77.5 | 369.5 | 303.6 | 483.6 | 86.8 |
| **2971** | 2971 | CHHATTISGARH | 1905 | 50.3 | 22.6 | 19.0 | 24.6 | 31.8 | 40.4 | 443.7 | 270.8 | 338.8 |
| **2972** | 2972 | CHHATTISGARH | 1906 | 25.0 | 91.0 | 52.5 | 0.0 | 4.1 | 210.1 | 445.2 | 258.3 | 242.3 |
| **2973** | 2973 | CHHATTISGARH | 1907 | 4.0 | 40.2 | 41.7 | 39.4 | 4.5 | 286.5 | 233.6 | 494.4 | 141.8 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **3077** | 3077 | CHHATTISGARH | 2011 | 0.3 | 11.5 | 2.6 | 35.0 | 16.8 | 183.5 | 272.6 | 379.8 | 382.2 |
| **3078** | 3078 | CHHATTISGARH | 2012 | 36.6 | 4.8 | 1.1 | 14.9 | 9.4 | 147.3 | 430.6 | 442.2 | 245.3 |
| **3079** | 3079 | CHHATTISGARH | 2013 | 2.8 | 19.7 | 4.9 | 45.8 | 5.7 | 263.6 | 418.8 | 336.6 | 140.9 |
| **3080** | 3080 | CHHATTISGARH | 2014 | 2.3 | 29.0 | 21.4 | 17.3 | 25.0 | 104.9 | 416.7 | 327.7 | 252.7 |
| **3081** | 3081 | CHHATTISGARH | 2015 | 15.8 | 1.2 | 21.2 | 37.0 | 13.0 | 257.6 | 248.6 | 286.6 | 216.9 |

113 rows × 20 columns

# Data Cleaning and Preprocessing

In [3]:

```
df.dropna()
```

Out[3]:

| | index | SUBDIVISION | YEAR | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2969** | 2969 | CHHATTISGARH | 1903 | 6.2 | 13.9 | 0.4 | 6.8 | 51.1 | 110.7 | 365.9 | 396.0 | 212.0 |
| **2970** | 2970 | CHHATTISGARH | 1904 | 0.0 | 8.6 | 32.3 | 0.2 | 77.5 | 369.5 | 303.6 | 483.6 | 86.8 |
| **2971** | 2971 | CHHATTISGARH | 1905 | 50.3 | 22.6 | 19.0 | 24.6 | 31.8 | 40.4 | 443.7 | 270.8 | 338.8 |
| **2972** | 2972 | CHHATTISGARH | 1906 | 25.0 | 91.0 | 52.5 | 0.0 | 4.1 | 210.1 | 445.2 | 258.3 | 242.3 |
| **2973** | 2973 | CHHATTISGARH | 1907 | 4.0 | 40.2 | 41.7 | 39.4 | 4.5 | 286.5 | 233.6 | 494.4 | 141.8 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **3077** | 3077 | CHHATTISGARH | 2011 | 0.3 | 11.5 | 2.6 | 35.0 | 16.8 | 183.5 | 272.6 | 379.8 | 382.2 |
| **3078** | 3078 | CHHATTISGARH | 2012 | 36.6 | 4.8 | 1.1 | 14.9 | 9.4 | 147.3 | 430.6 | 442.2 | 245.3 |
| **3079** | 3079 | CHHATTISGARH | 2013 | 2.8 | 19.7 | 4.9 | 45.8 | 5.7 | 263.6 | 418.8 | 336.6 | 140.9 |
| **3080** | 3080 | CHHATTISGARH | 2014 | 2.3 | 29.0 | 21.4 | 17.3 | 25.0 | 104.9 | 416.7 | 327.7 | 252.7 |
| **3081** | 3081 | CHHATTISGARH | 2015 | 15.8 | 1.2 | 21.2 | 37.0 | 13.0 | 257.6 | 248.6 | 286.6 | 216.9 |

113 rows × 20 columns

In [4]:

```
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, 2969 to 3081
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
```

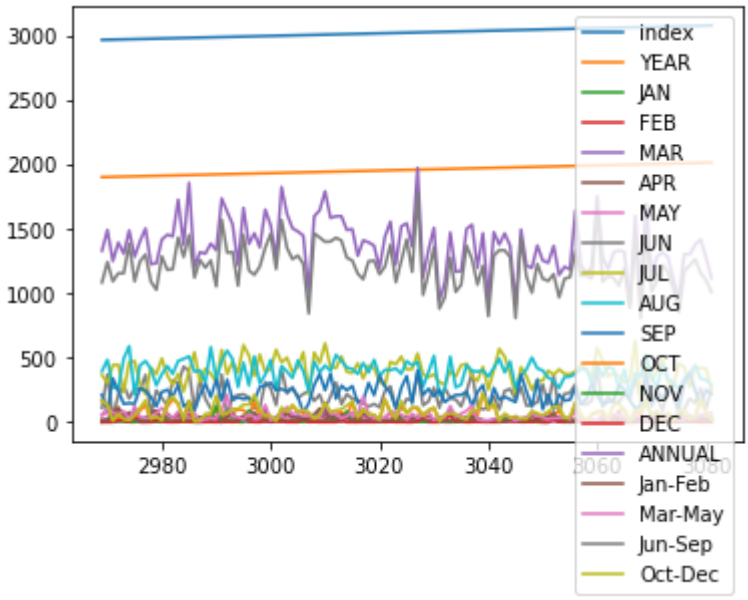
# Line Chart

In [6]:

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

Out[6]:

<AxesSubplot:>



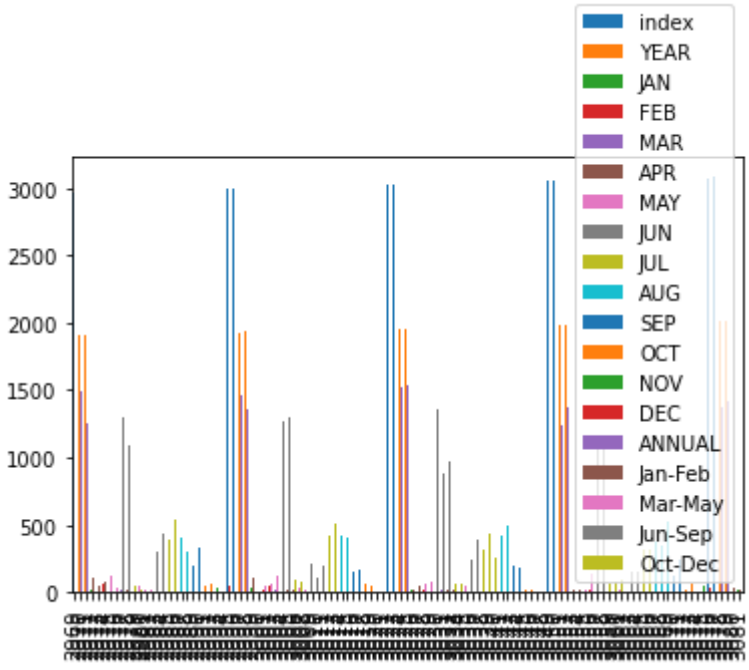

# Bar chart

In [7]:

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

Out[7]:

<AxesSubplot:>



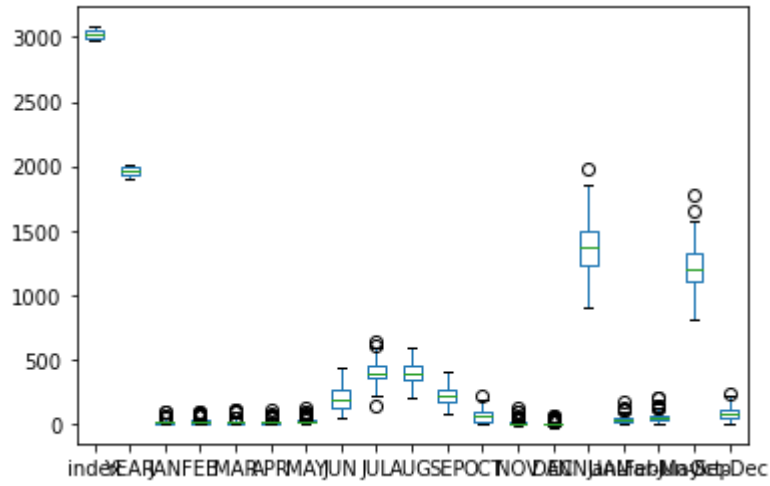
# Box chart

In [8]:

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

Out[8]:

<AxesSubplot:>



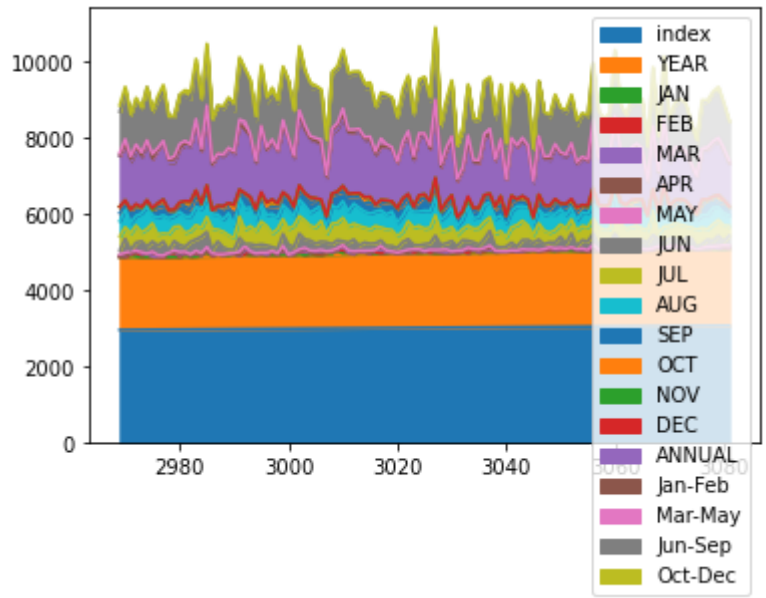

# Area Chart

In [9]:

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

Out[9]:

<AxesSubplot:>



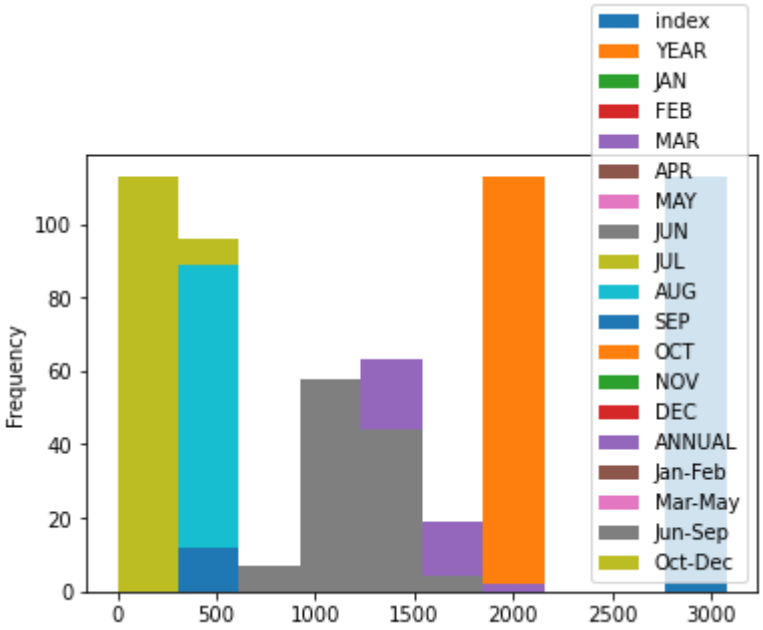

# Histogram

In [10]:

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

Out[10]:

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




# pie chart

In [11]:

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

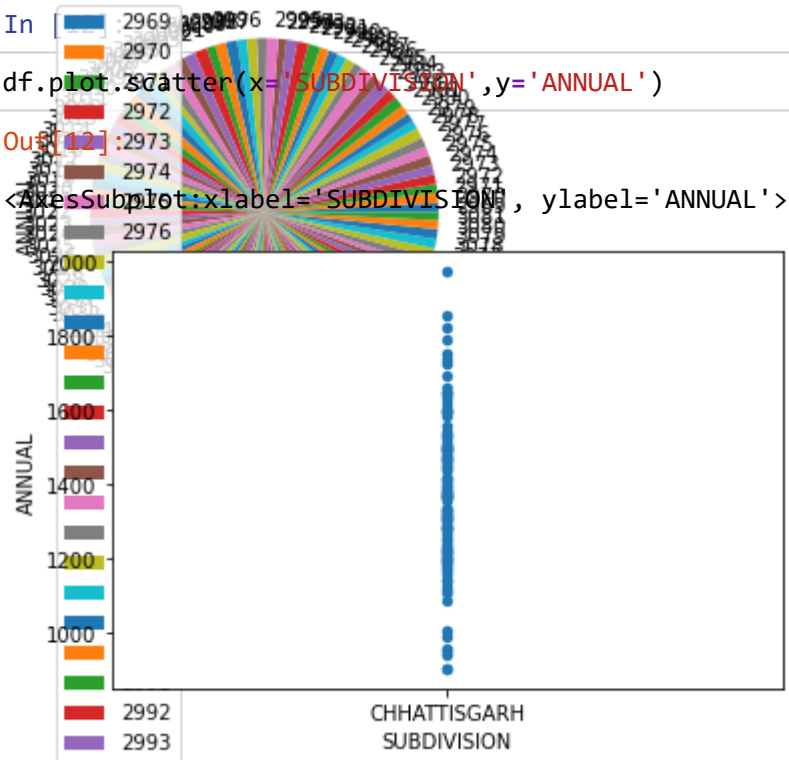
Out[11]:

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

# Scatter chart

In [ ]:
```
df.plot.scatter(x='SUBDIVISION',y='ANNUAL')
```

Out[12]:
```
<AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>
```




In [ ]:
```
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 | 1 |
| **mean** | 3025.00000 | 1959.00000 | 14.020354 | 18.511504 | 15.287611 | 16.898230 | 21.167257 | 2 |
| **std** | 32.76431 | 32.76431 | 18.138998 | 21.301709 | 20.756342 | 17.241841 | 20.787090 | |
| **min** | 2969.00000 | 1903.00000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| **25%** | 2997.00000 | 1931.00000 | 1.700000 | 4.000000 | 2.200000 | 5.200000 | 8.700000 | 1 |
| **50%** | 3025.00000 | 1959.00000 | 6.500000 | 11.000000 | 7.900000 | 11.900000 | 15.500000 | 1 |
| **75%** | 3053.00000 | 1987.00000 | 21.100000 | 24.800000 | 21.200000 | 21.200000 | 25.300000 | 2 |
| **max** | 3081.00000 | 2015.00000 | 99.500000 | 91.000000 | 102.900000 | 112.800000 | 122.300000 | 4 |

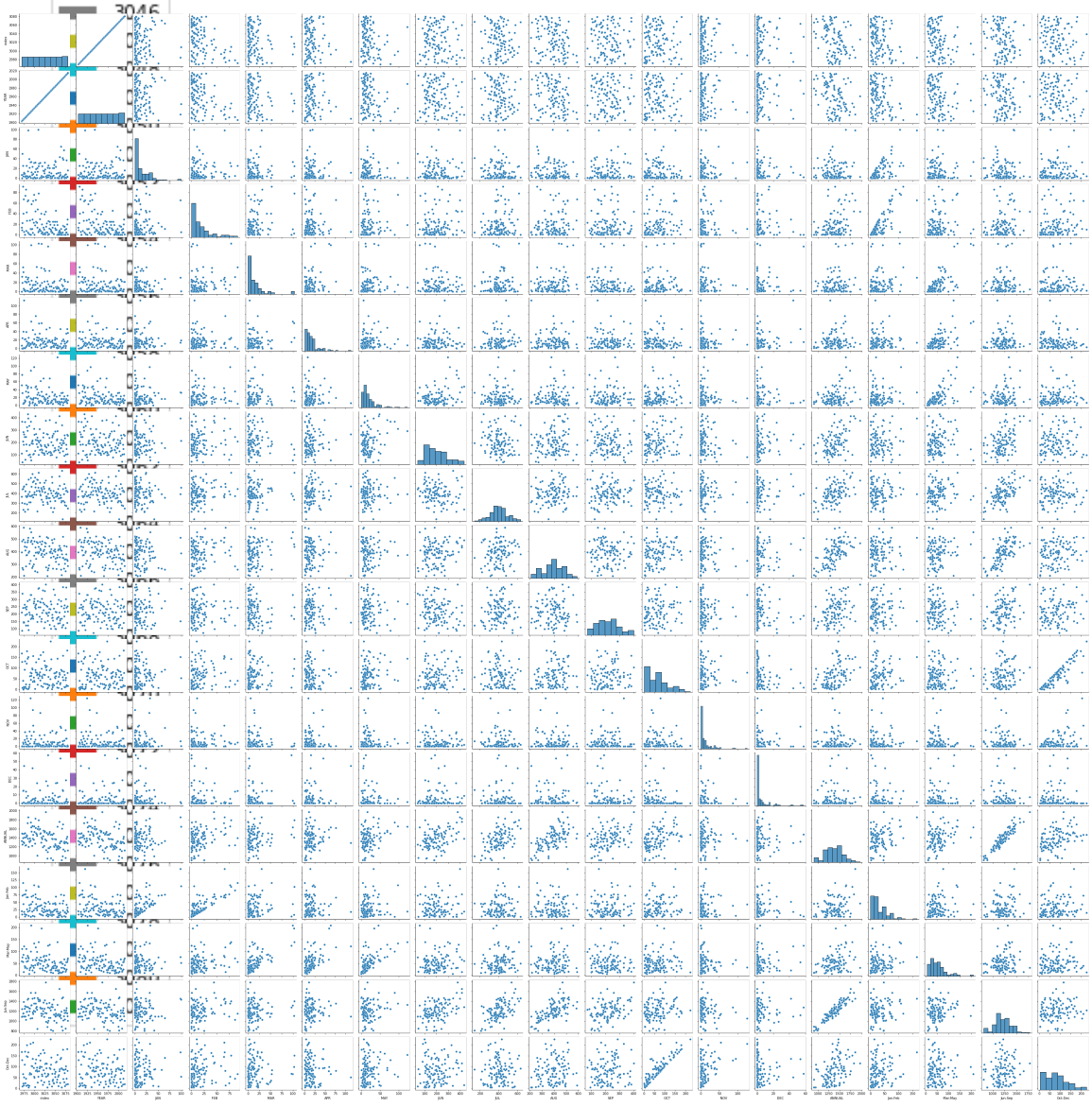
# EDA and Visualization

In [14]:

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

Out[14]:

```
<seaborn.axisgrid.PairGrid at 0x1c96220b430>
```

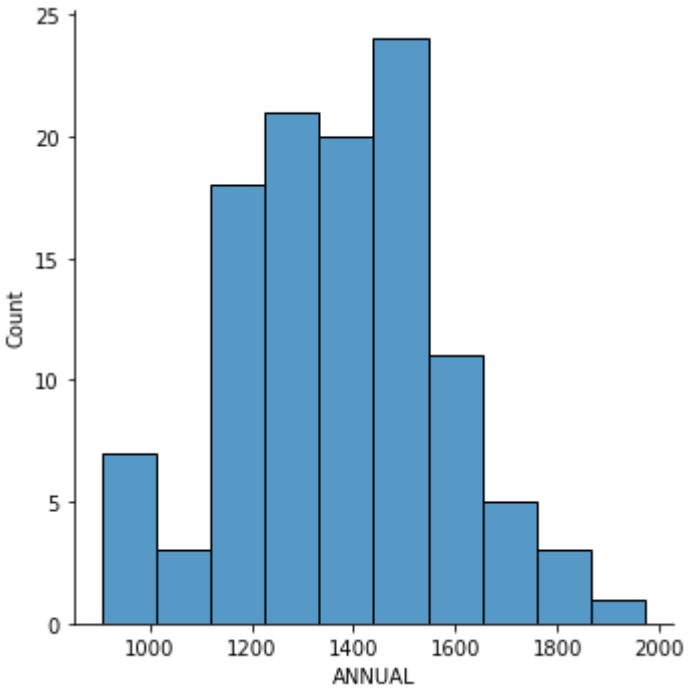


In [15]:

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

Out[15]:

```
<seaborn.axisgrid.FacetGrid at 0x1c97069b640>
```

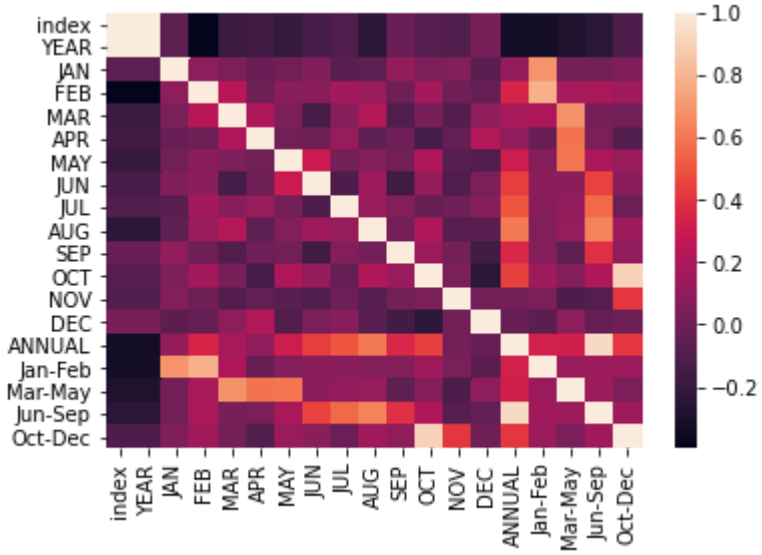



In [16]:

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

Out[16]:

```
<AxesSubplot:>
```




In [ ]: