

# Import Libraries

In [1]:

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

In [11]:

```
d=pd.read_csv(r"C:\Users\user\Downloads\FP2_RainFall\rain.csv")[1357:1472]
d
```

Out[11]:

|      | index | SUBDIVISION                      | YEAR | JAN  | FEB  | MAR  | APR  | MAY  | JUN  | JUL   | AUG   | SEP   | C |
|------|-------|----------------------------------|------|------|------|------|------|------|------|-------|-------|-------|---|
| 1357 | 1357  | HARYANA<br>DELHI &<br>CHANDIGARH | 1901 | 35.4 | 28.9 | 11.1 | 0.0  | 5.1  | 13.2 | 126.4 | 151.5 | 10.5  |   |
| 1358 | 1358  | HARYANA<br>DELHI &<br>CHANDIGARH | 1902 | 0.0  | 0.7  | 2.9  | 10.2 | 15.8 | 74.6 | 149.3 | 97.1  | 59.8  |   |
| 1359 | 1359  | HARYANA<br>DELHI &<br>CHANDIGARH | 1903 | 14.7 | 0.5  | 2.3  | 0.5  | 8.5  | 8.6  | 151.6 | 138.2 | 97.7  |   |
| 1360 | 1360  | HARYANA<br>DELHI &<br>CHANDIGARH | 1904 | 7.6  | 0.7  | 48.0 | 0.5  | 29.3 | 34.3 | 109.7 | 162.9 | 102.3 |   |
| 1361 | 1361  | HARYANA<br>DELHI &<br>CHANDIGARH | 1905 | 44.8 | 20.8 | 14.0 | 1.3  | 7.4  | 20.1 | 93.6  | 23.1  | 92.6  |   |
| ...  | ...   | ...                              | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...   | ...   | ...   |   |
| 1467 | 1467  | HARYANA<br>DELHI &<br>CHANDIGARH | 2011 | 0.7  | 26.7 | 6.9  | 8.9  | 28.7 | 94.4 | 85.0  | 127.3 | 133.1 |   |
| 1468 | 1468  | HARYANA<br>DELHI &<br>CHANDIGARH | 2012 | 8.2  | 0.2  | 0.1  | 11.8 | 3.8  | 5.3  | 68.1  | 196.6 | 90.7  |   |
| 1469 | 1469  | HARYANA<br>DELHI &<br>CHANDIGARH | 2013 | 21.1 | 52.2 | 5.3  | 3.3  | 1.4  | 62.1 | 96.5  | 161.9 | 42.8  | 1 |
| 1470 | 1470  | HARYANA<br>DELHI &<br>CHANDIGARH | 2014 | 13.0 | 17.3 | 26.8 | 7.5  | 20.3 | 25.9 | 72.3  | 34.8  | 67.3  | 1 |
| 1471 | 1471  | HARYANA<br>DELHI &<br>CHANDIGARH | 2015 | 12.4 | 6.6  | 71.8 | 34.8 | 8.4  | 43.7 | 130.3 | 89.2  | 32.1  |   |

115 rows × 20 columns



# Data Cleaning and preprocessing

In [13]:

```
d.dropna()
```

Out[13]:

|      | index | SUBDIVISION                      | YEAR | JAN  | FEB  | MAR  | APR  | MAY  | JUN  | JUL   | AUG   | SEP   | C |
|------|-------|----------------------------------|------|------|------|------|------|------|------|-------|-------|-------|---|
| 1357 | 1357  | HARYANA<br>DELHI &<br>CHANDIGARH | 1901 | 35.4 | 28.9 | 11.1 | 0.0  | 5.1  | 13.2 | 126.4 | 151.5 | 10.5  |   |
| 1358 | 1358  | HARYANA<br>DELHI &<br>CHANDIGARH | 1902 | 0.0  | 0.7  | 2.9  | 10.2 | 15.8 | 74.6 | 149.3 | 97.1  | 59.8  |   |
| 1359 | 1359  | HARYANA<br>DELHI &<br>CHANDIGARH | 1903 | 14.7 | 0.5  | 2.3  | 0.5  | 8.5  | 8.6  | 151.6 | 138.2 | 97.7  |   |
| 1360 | 1360  | HARYANA<br>DELHI &<br>CHANDIGARH | 1904 | 7.6  | 0.7  | 48.0 | 0.5  | 29.3 | 34.3 | 109.7 | 162.9 | 102.3 |   |
| 1361 | 1361  | HARYANA<br>DELHI &<br>CHANDIGARH | 1905 | 44.8 | 20.8 | 14.0 | 1.3  | 7.4  | 20.1 | 93.6  | 23.1  | 92.6  |   |
| ...  | ...   | ...                              | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...   | ...   | ...   |   |
| 1467 | 1467  | HARYANA<br>DELHI &<br>CHANDIGARH | 2011 | 0.7  | 26.7 | 6.9  | 8.9  | 28.7 | 94.4 | 85.0  | 127.3 | 133.1 |   |
| 1468 | 1468  | HARYANA<br>DELHI &<br>CHANDIGARH | 2012 | 8.2  | 0.2  | 0.1  | 11.8 | 3.8  | 5.3  | 68.1  | 196.6 | 90.7  |   |
| 1469 | 1469  | HARYANA<br>DELHI &<br>CHANDIGARH | 2013 | 21.1 | 52.2 | 5.3  | 3.3  | 1.4  | 62.1 | 96.5  | 161.9 | 42.8  | 1 |
| 1470 | 1470  | HARYANA<br>DELHI &<br>CHANDIGARH | 2014 | 13.0 | 17.3 | 26.8 | 7.5  | 20.3 | 25.9 | 72.3  | 34.8  | 67.3  | 1 |
| 1471 | 1471  | HARYANA<br>DELHI &<br>CHANDIGARH | 2015 | 12.4 | 6.6  | 71.8 | 34.8 | 8.4  | 43.7 | 130.3 | 89.2  | 32.1  |   |

115 rows × 20 columns

In [14]:

```
d.columns
```

Out[14]:

```
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 [15]:

```
d.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 115 entries, 1357 to 1471
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  -
0   index                 115 non-null    int64
1   SUBDIVISION           115 non-null    object
2   YEAR                  115 non-null    int64
3   JAN                   115 non-null    float64
4   FEB                   115 non-null    float64
5   MAR                   115 non-null    float64
6   APR                   115 non-null    float64
7   MAY                   115 non-null    float64
8   JUN                   115 non-null    float64
9   JUL                   115 non-null    float64
10  AUG                   115 non-null    float64
11  SEP                   115 non-null    float64
12  OCT                   115 non-null    float64
13  NOV                   115 non-null    float64
14  DEC                   115 non-null    float64
15  ANNUAL                115 non-null    float64
16  Jan-Feb               115 non-null    float64
17  Mar-May               115 non-null    float64
18  Jun-Sep               115 non-null    float64
19  Oct-Dec               115 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 18.1+ KB
```

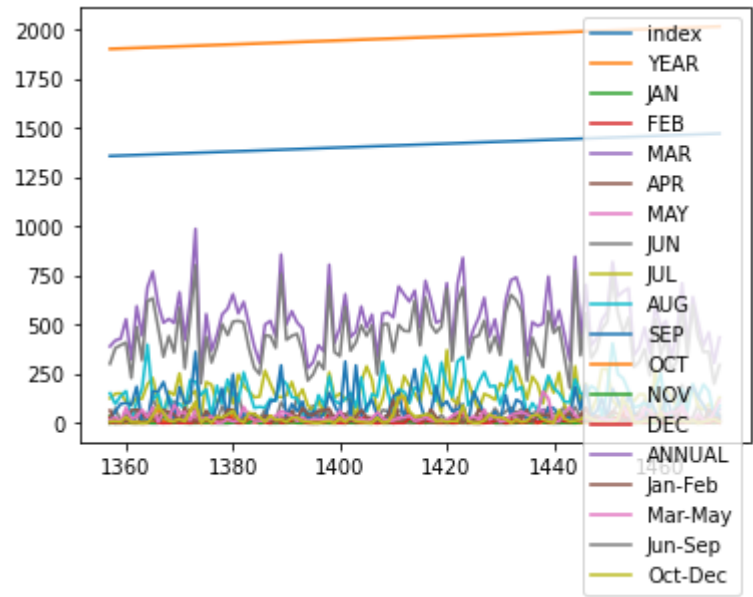
# Line Chart

In [16]:

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

Out[16]:

<AxesSubplot:>



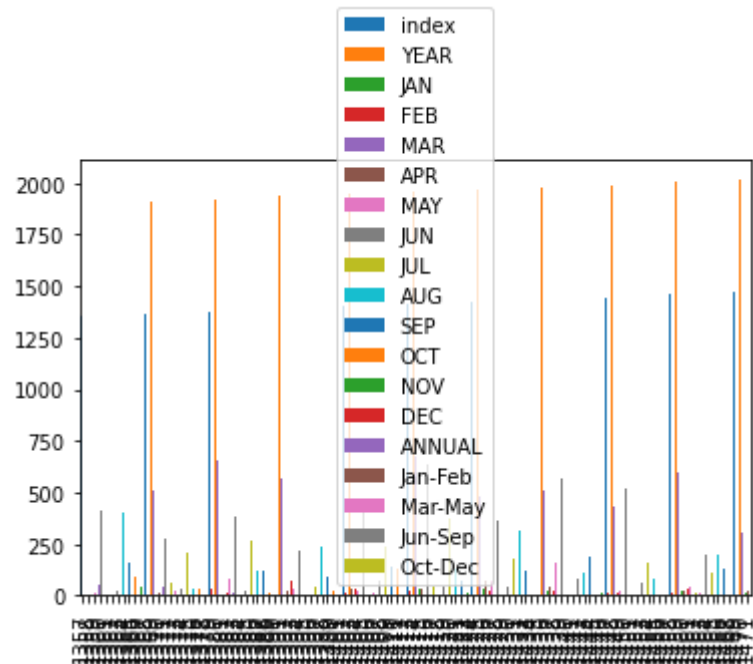
# Bar Chart

In [17]:

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

Out[17]:

<AxesSubplot:>



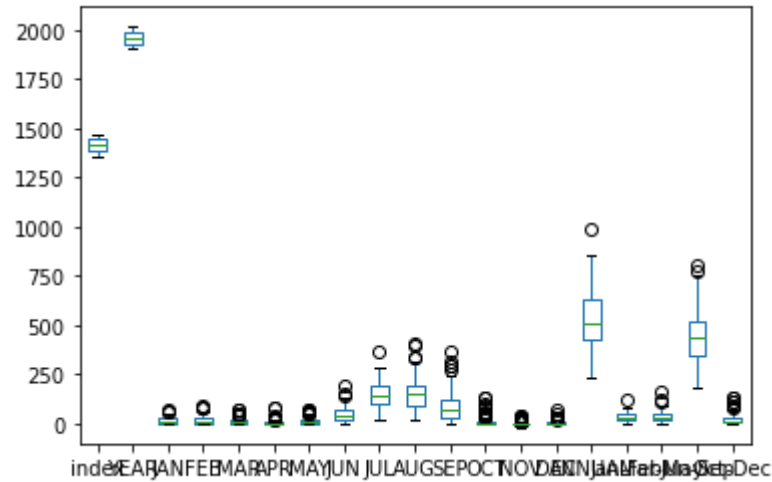
# Box Chart

In [18]:

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

Out[18]:

<AxesSubplot:>



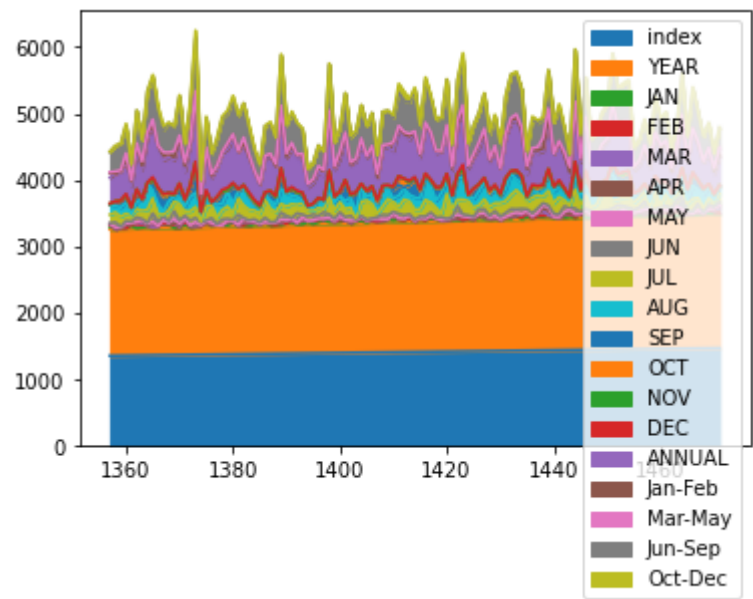
# Area Chart

In [19]:

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

Out[19]:

<AxesSubplot:>



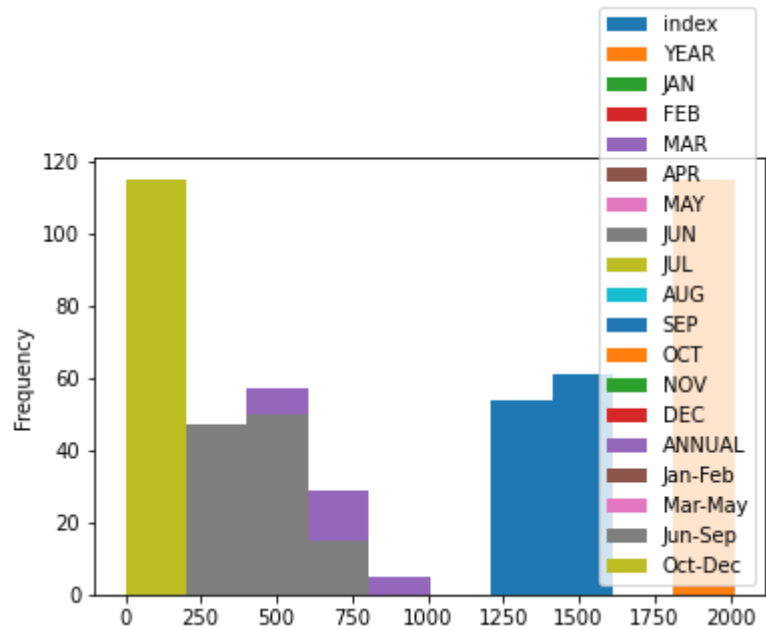
# Histogram

In [20]:

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

Out[20]:

<AxesSubplot:ylabel='Frequency'>



# Pie Chart

In [21]:

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

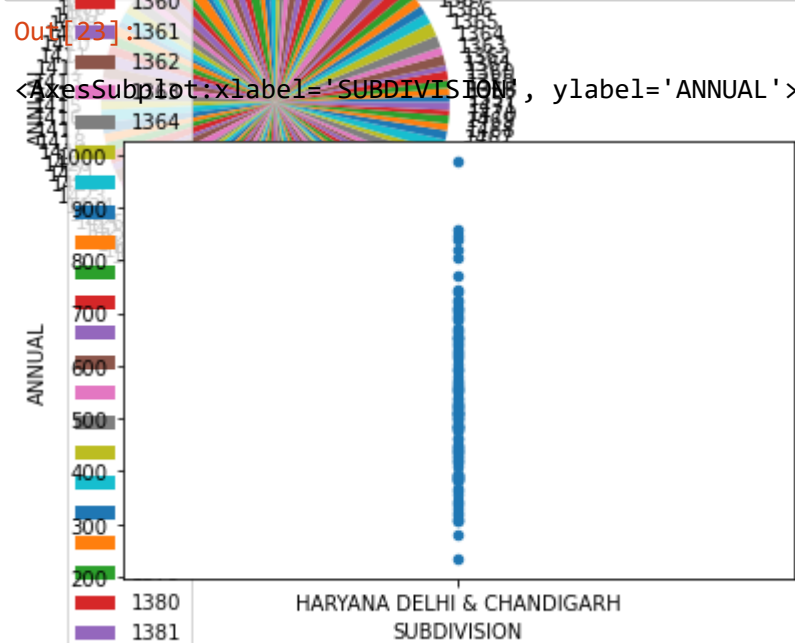
Out[21]:

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

## Scatter Chart



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



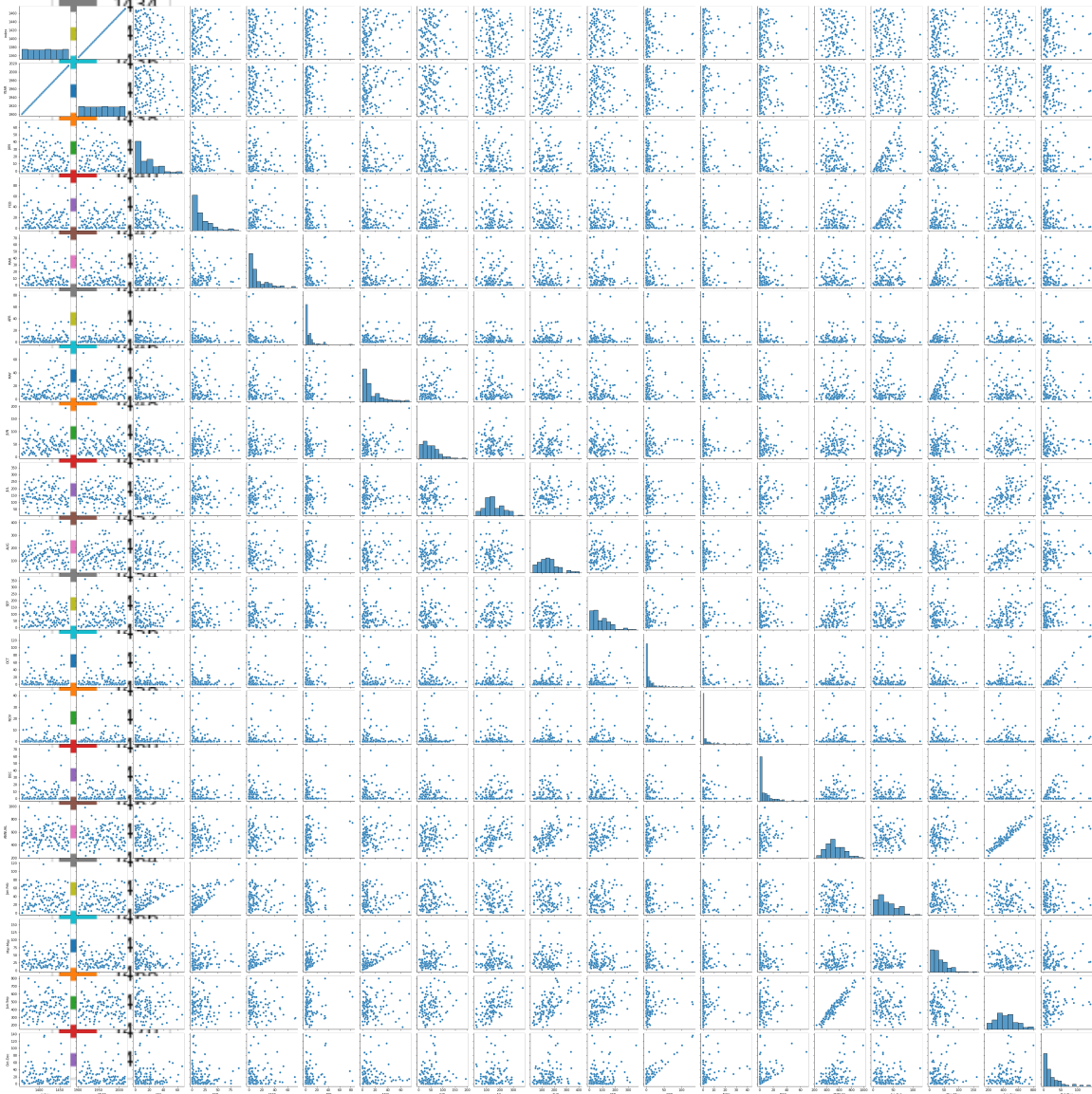
```
In [ ]: d.describe()
```

Out[24]:

|       | index       | YEAR        | JAN        | FEB        | MAR        | APR        | MAY        |
|-------|-------------|-------------|------------|------------|------------|------------|------------|
| count | 115.000000  | 115.000000  | 115.000000 | 115.000000 | 115.000000 | 115.000000 | 115.000000 |
| mean  | 1414.000000 | 1958.000000 | 16.889565  | 17.433913  | 12.935652  | 7.633913   | 14.533913  |
| std   | 33.341666   | 33.341666   | 15.514478  | 18.893422  | 15.251840  | 12.847533  | 15.900347  |
| min   | 1357.000000 | 1901.000000 | 0.000000   | 0.000000   | 0.000000   | 0.000000   | 0.000000   |
| 25%   | 1385.500000 | 1929.500000 | 3.550000   | 2.250000   | 2.100000   | 0.800000   | 3.700000   |
| 50%   | 1414.000000 | 1958.000000 | 14.300000  | 12.100000  | 7.200000   | 2.800000   | 7.900000   |
| 75%   | 1442.500000 | 1986.500000 | 25.150000  | 27.850000  | 17.700000  | 8.750000   | 20.700000  |
| max   | 1471.000000 | 2015.000000 | 66.500000  | 91.000000  | 71.800000  | 82.500000  | 72.900000  |

# EDA AND VISUALIZATION

```
In [ ]: sns.pairplot(d)
Out[25]: <seaborn.axisgrid.PairGrid at 0x1e3381f2e80>
```

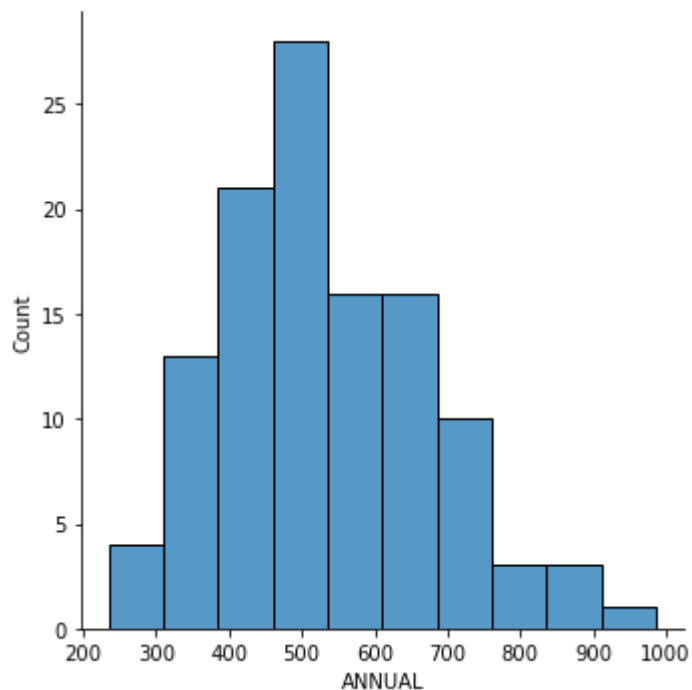


In [27]:

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

Out[27]:

```
<seaborn.axisgrid.FacetGrid at 0x1e342047b50>
```

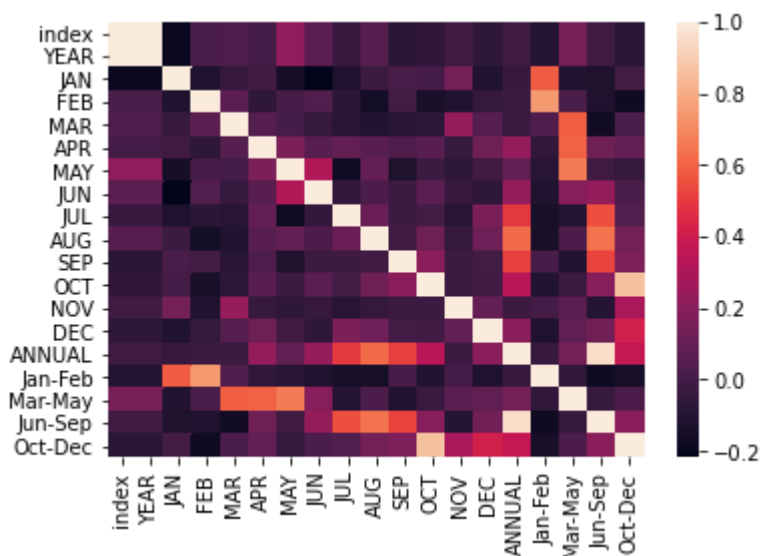


In [28]:

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

Out[28]:

```
<AxesSubplot:>
```



In [ ]:

