Import Libraies

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 in india 1901-2015.csv")[1817:193
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
1817	1817	WEST RAJASTHAN	1901	6.7	0.0	1.1	0.0	6.1	3.0	79.0	59.2	1.0	2.1	0.0	0.6	158.9
1818	1818	WEST RAJASTHAN	1902	0.0	0.0	0.0	0.5	4.0	49.1	27.0	71.3	41.8	1.8	0.0	0.0	195.6
1819	1819	WEST RAJASTHAN	1903	1.7	1.3	5.5	0.0	4.2	2.7	154.8	87.1	49.3	0.1	0.0	0.5	307.0
1820	1820	WEST RAJASTHAN	1904	3.8	2.9	16.3	0.7	11.4	14.6	39.8	45.6	21.4	1.4	2.9	7.1	167.9
1821	1821	WEST RAJASTHAN	1905	6.3	4.8	0.7	1.3	0.3	4.9	30.1	0.6	64.5	0.0	0.0	0.9	114.4
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1926	1926	WEST RAJASTHAN	2010	1.9	1.7	0.6	0.6	3.0	49.5	129.9	163.4	96.5	0.9	20.1	5.2	473.2
1927	1927	WEST RAJASTHAN	2011	0.0	11.8	1.5	1.5	7.8	24.4	88.5	166.8	116.3	0.1	0.0	0.0	418.7
1928	1928	WEST RAJASTHAN	2012	0.5	0.0	0.0	9.5	10.4	5.3	40.4	166.7	92.0	1.9	0.0	0.6	327.3
1929	1929	WEST RAJASTHAN	2013	8.6	21.8	4.2	3.1	1.7	37.6	104.5	138.2	58.7	10.1	1.0	0.0	389.4
1930	1930	WEST RAJASTHAN	2014	0.8	2.2	4.7	8.4	23.0	13.8	94.3	69.6	84.9	0.5	0.2	0.0	302.4

114 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	OCT	NOV	DEC	ANNUAL
1817	1817	WEST RAJASTHAN	1901	6.7	0.0	1.1	0.0	6.1	3.0	79.0	59.2	1.0	2.1	0.0	0.6	158.9
1818	1818	WEST RAJASTHAN	1902	0.0	0.0	0.0	0.5	4.0	49.1	27.0	71.3	41.8	1.8	0.0	0.0	195.6
1819	1819	WEST RAJASTHAN	1903	1.7	1.3	5.5	0.0	4.2	2.7	154.8	87.1	49.3	0.1	0.0	0.5	307.0
1820	1820	WEST RAJASTHAN	1904	3.8	2.9	16.3	0.7	11.4	14.6	39.8	45.6	21.4	1.4	2.9	7.1	167.9
1821	1821	WEST RAJASTHAN	1905	6.3	4.8	0.7	1.3	0.3	4.9	30.1	0.6	64.5	0.0	0.0	0.9	114.4
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1926	1926	WEST RAJASTHAN	2010	1.9	1.7	0.6	0.6	3.0	49.5	129.9	163.4	96.5	0.9	20.1	5.2	473.2
1927	1927	WEST RAJASTHAN	2011	0.0	11.8	1.5	1.5	7.8	24.4	88.5	166.8	116.3	0.1	0.0	0.0	418.7
1928	1928	WEST RAJASTHAN	2012	0.5	0.0	0.0	9.5	10.4	5.3	40.4	166.7	92.0	1.9	0.0	0.6	327.3
1929	1929	WEST RAJASTHAN	2013	8.6	21.8	4.2	3.1	1.7	37.6	104.5	138.2	58.7	10.1	1.0	0.0	389.4
1930	1930	WEST RAJASTHAN	2014	0.8	2.2	4.7	8.4	23.0	13.8	94.3	69.6	84.9	0.5	0.2	0.0	302.4

114 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-Feb',
       'Mar-May', 'Jun-Sep', 'Oct-Dec'],
      dtype='object')
```

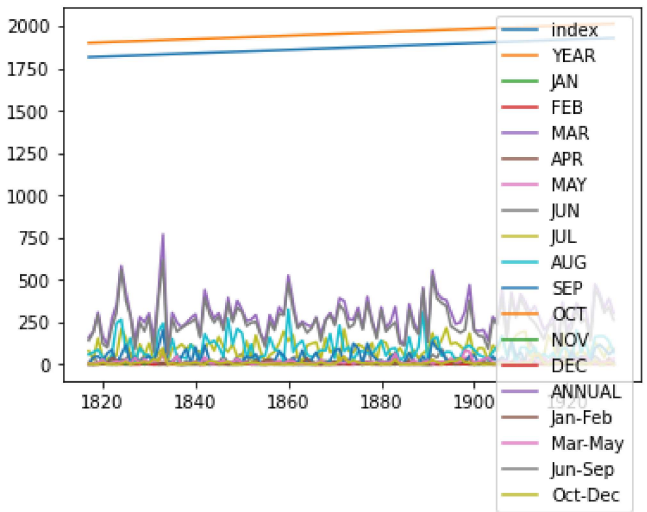
In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 114 entries, 1817 to 1930
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]:
```
df.plot.line()
```

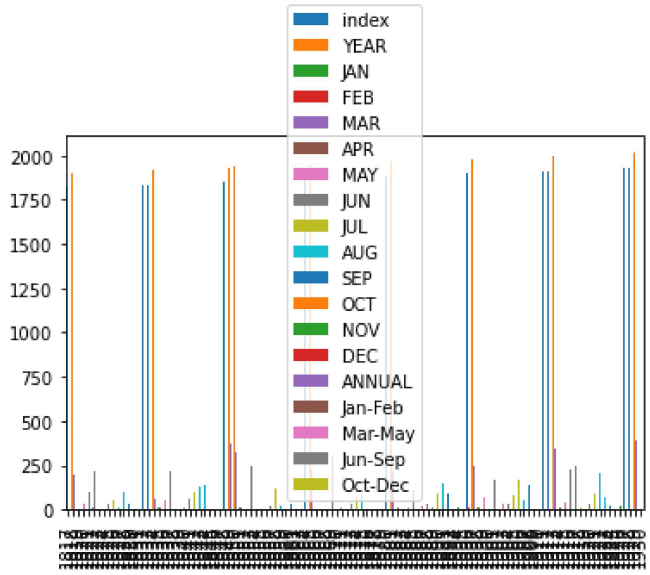
Out[6]: <AxesSubplot:>



Bar chart

In [7]:
```
df.plot.bar()
```

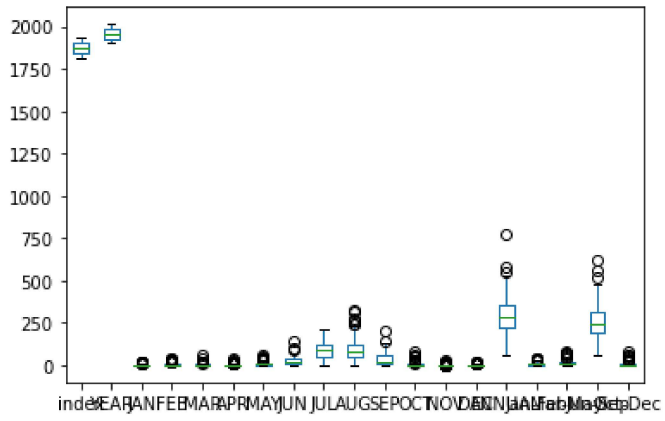
Out[7]: <AxesSubplot:>



Box chart

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

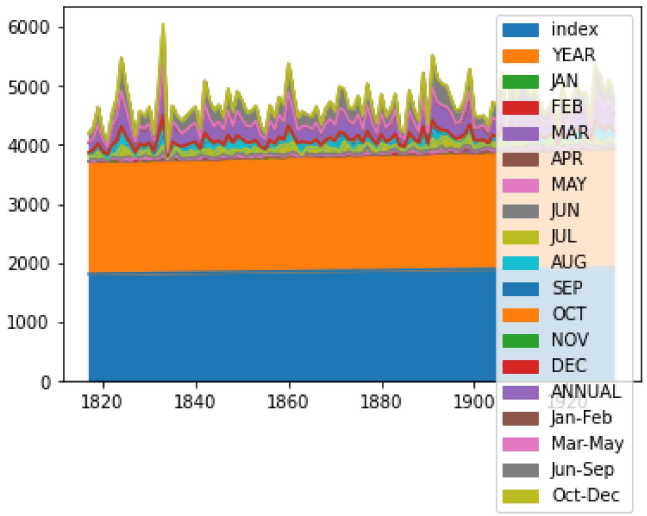
Out[8]: <AxesSubplot:>



Area chart

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

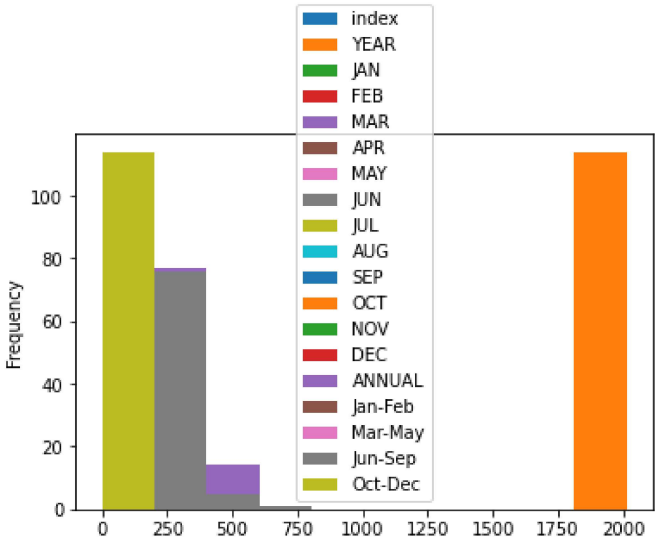
Out[9]: <AxesSubplot:>



Histogram

In [10]:
```
df.plot.hist()
```

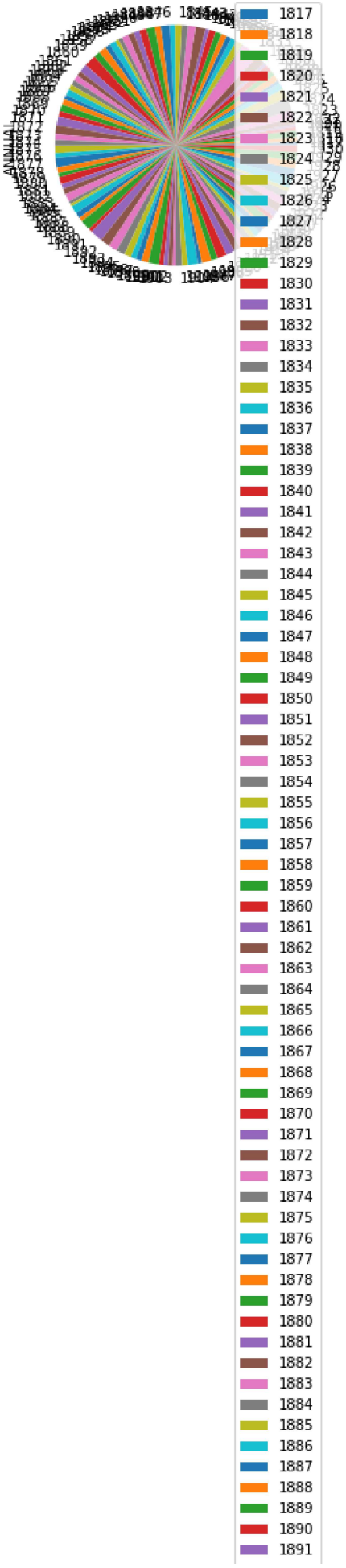
Out[10]: <AxesSubplot:ylabel='Frequency'>

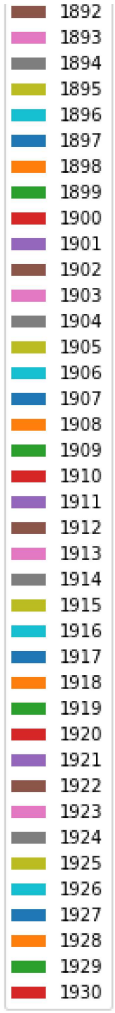


pie chart

In [11]:
```
df.plot.pie(y="ANNUAL")
```

Out[11]:
```
<AxesSubplot:ylabel='ANNUAL'>
```

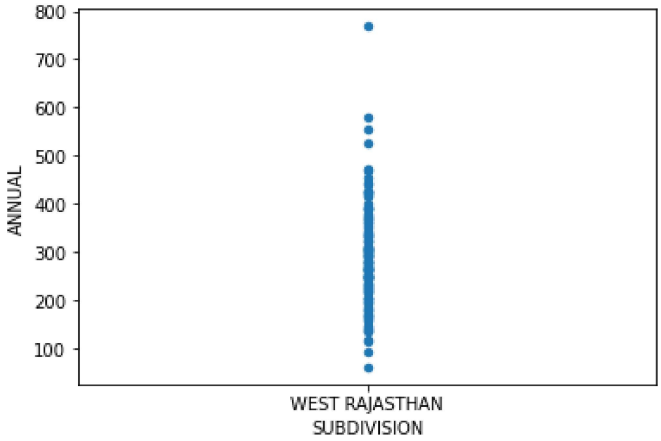

1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891



Scatter chart

In [12]:
```
df.plot.scatter(y='ANNUAL',x='SUBDIVISION')
```

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



In [13]:
```
df.describe()
```

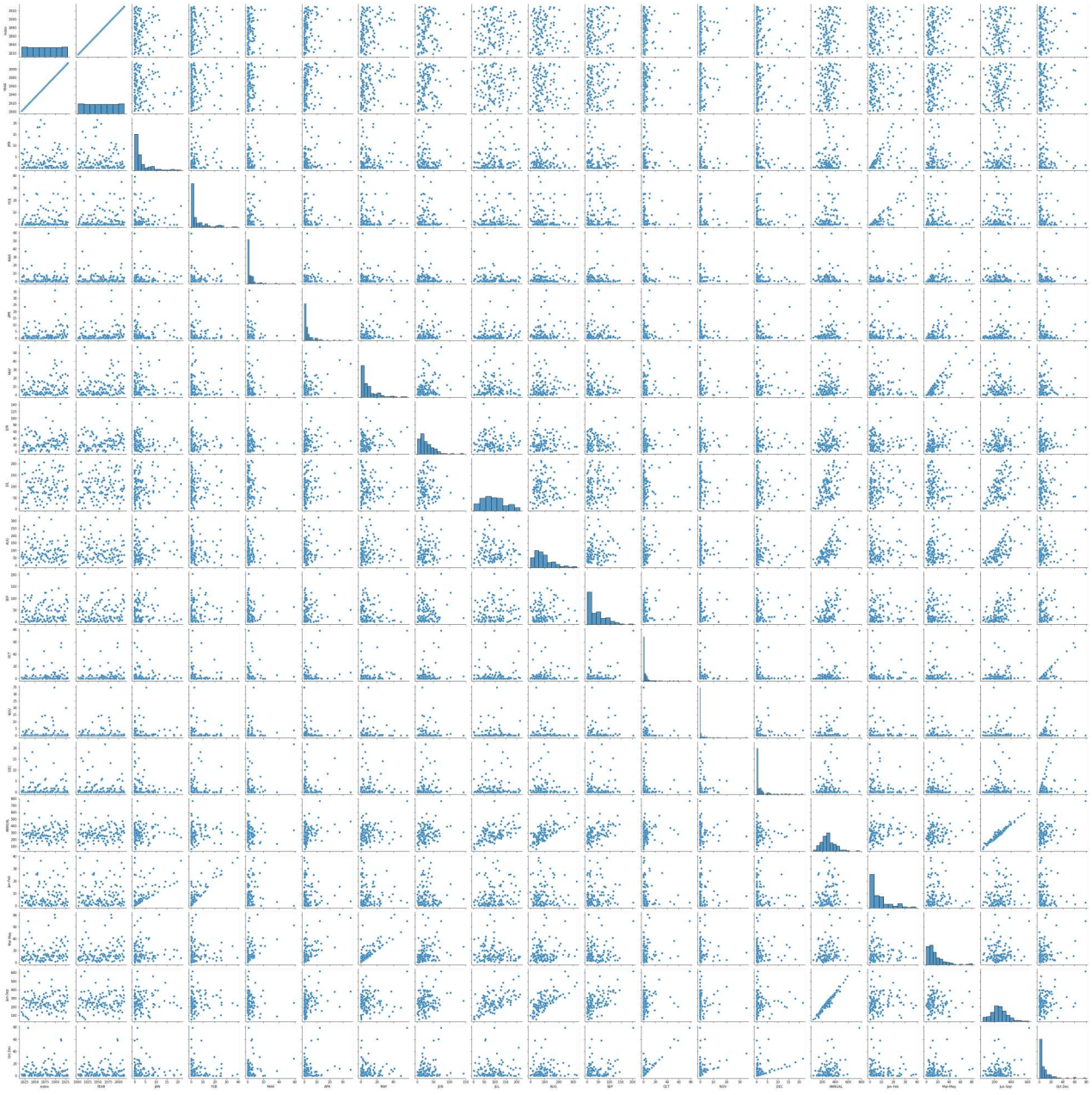
Out[13]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL
count	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000
mean	1873.500000	1957.500000	3.344737	4.965789	3.755263	3.381579	9.390351	28.421930	93.948246
std	33.052988	33.052988	4.568381	7.884305	7.444260	5.580522	10.886055	22.792216	51.189478
min	1817.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.400000	2.400000
25%	1845.250000	1929.250000	0.325000	0.200000	0.200000	0.400000	1.925000	13.050000	54.025000
50%	1873.500000	1957.500000	1.650000	1.300000	1.100000	1.400000	6.050000	21.550000	89.500000
75%	1901.750000	1985.750000	4.100000	6.075000	5.000000	3.675000	12.000000	39.100000	125.975000
max	1930.000000	2014.000000	21.400000	39.100000	59.000000	36.100000	56.800000	143.200000	215.400000

EDA AND VISUALIZATION

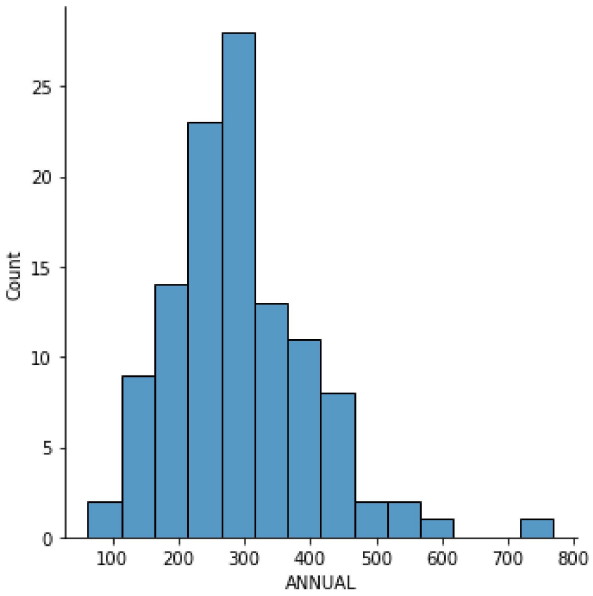
In [14]:
```
sns.pairplot(df)
```

Out[14]:
```
<seaborn.axisgrid.PairGrid at 0x206ac2f1670>
```



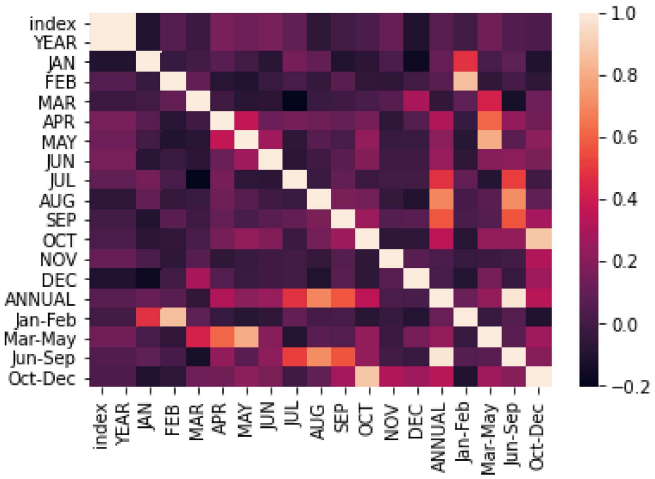
In [15]:
```
sns.displot(df['ANNUAL'])
```

Out[15]: <seaborn.axisgrid.FacetGrid at 0x206b16a1a30>



In [16]:
```
sns.heatmap(df.corr())
```

Out[16]: <AxesSubplot:>



In [ ]:
```

```