Import Libraies

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [3]:
```
df=pd.read_csv(r"C:\Users\user\Downloads\FP2_RainFall\rainfall in india 1901-2015.csv")[1474:158
df
```

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
1474	1474	PUNJAB	1903	29.5	0.5	45.0	1.3	9.2	5.2	212.2	119.1	132.5	6.9	0.0	9.5	571.0
1475	1475	PUNJAB	1904	24.2	1.7	87.8	1.2	13.8	22.0	59.9	124.0	73.8	7.4	9.8	25.9	451.5
1476	1476	PUNJAB	1905	53.0	40.3	24.3	0.5	2.2	19.2	122.6	50.3	111.1	1.2	0.0	9.4	434.3
1477	1477	PUNJAB	1906	5.3	83.1	50.5	1.5	4.7	38.3	99.4	190.5	181.5	0.5	0.0	22.2	677.5
1478	1478	PUNJAB	1907	27.3	95.4	53.1	50.7	3.3	20.2	82.8	249.0	6.8	0.4	0.0	0.0	589.2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1582	1582	PUNJAB	2011	3.5	35.6	8.2	17.8	18.9	162.9	120.9	193.5	140.2	0.0	1.0	2.6	705.2
1583	1583	PUNJAB	2012	62.6	3.2	1.9	31.1	1.6	11.9	120.2	135.1	112.3	2.2	0.4	11.0	493.6
1584	1584	PUNJAB	2013	9.3	50.1	11.6	3.4	3.6	120.3	117.9	217.1	24.4	16.2	6.1	6.6	586.6
1585	1585	PUNJAB	2014	21.8	20.1	30.3	24.5	20.8	20.6	76.3	41.9	105.8	6.0	0.7	14.1	382.7
1586	1586	PUNJAB	2015	17.7	31.3	68.5	29.8	16.7	48.3	130.2	88.6	69.2	9.0	0.8	0.7	510.8

113 rows × 20 columns

Data Cleaning and Preprocessing

In [4]:
```
df.dropna()
```

Out[4]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
1474	1474	PUNJAB	1903	29.5	0.5	45.0	1.3	9.2	5.2	212.2	119.1	132.5	6.9	0.0	9.5	571.0
1475	1475	PUNJAB	1904	24.2	1.7	87.8	1.2	13.8	22.0	59.9	124.0	73.8	7.4	9.8	25.9	451.5
1476	1476	PUNJAB	1905	53.0	40.3	24.3	0.5	2.2	19.2	122.6	50.3	111.1	1.2	0.0	9.4	434.3
1477	1477	PUNJAB	1906	5.3	83.1	50.5	1.5	4.7	38.3	99.4	190.5	181.5	0.5	0.0	22.2	677.5
1478	1478	PUNJAB	1907	27.3	95.4	53.1	50.7	3.3	20.2	82.8	249.0	6.8	0.4	0.0	0.0	589.2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1582	1582	PUNJAB	2011	3.5	35.6	8.2	17.8	18.9	162.9	120.9	193.5	140.2	0.0	1.0	2.6	705.2
1583	1583	PUNJAB	2012	62.6	3.2	1.9	31.1	1.6	11.9	120.2	135.1	112.3	2.2	0.4	11.0	493.6
1584	1584	PUNJAB	2013	9.3	50.1	11.6	3.4	3.6	120.3	117.9	217.1	24.4	16.2	6.1	6.6	586.6
1585	1585	PUNJAB	2014	21.8	20.1	30.3	24.5	20.8	20.6	76.3	41.9	105.8	6.0	0.7	14.1	382.7
1586	1586	PUNJAB	2015	17.7	31.3	68.5	29.8	16.7	48.3	130.2	88.6	69.2	9.0	0.8	0.7	510.8

113 rows × 20 columns

In [5]:
```
df.columns
```

Out[5]:
```
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')
```

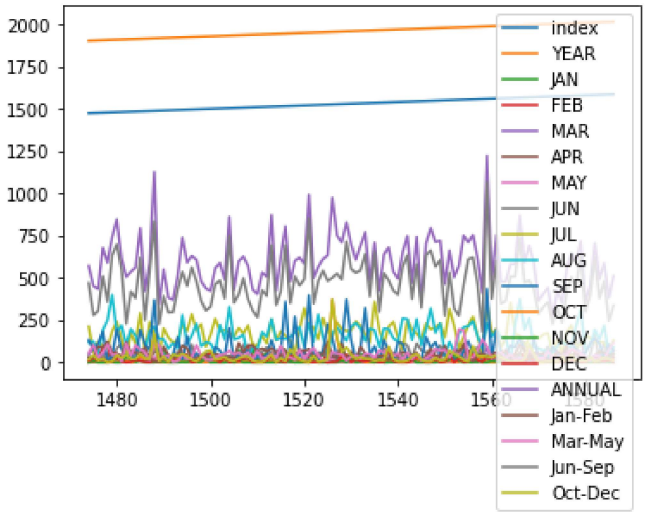
In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113 entries, 1474 to 1586
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 [7]:
```
df.plot.line()
```

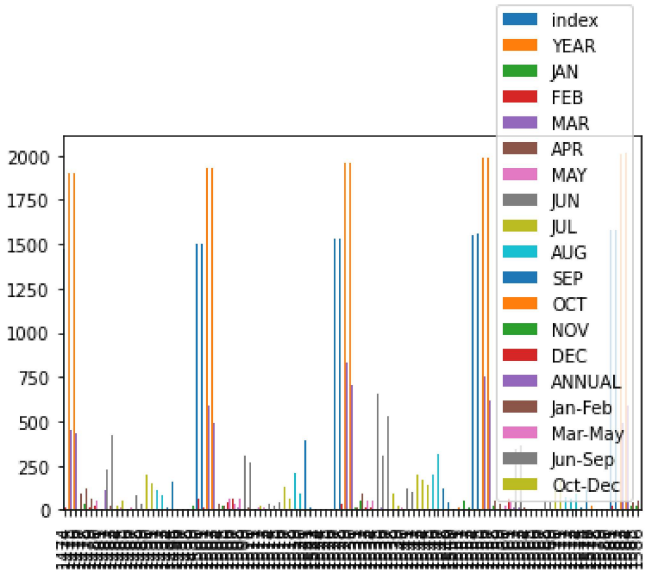
Out[7]: <AxesSubplot:>



Bar chart

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

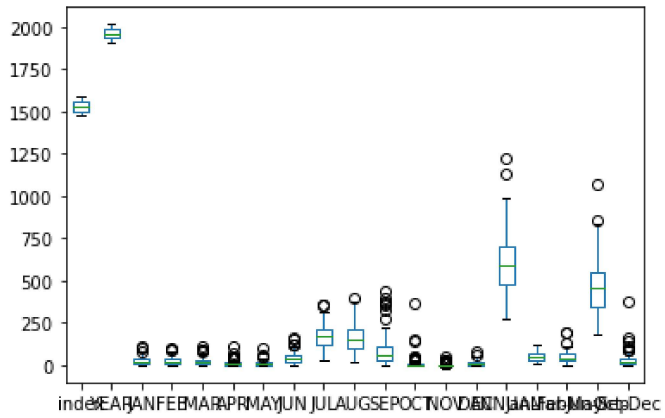
Out[9]: <AxesSubplot:>



Box chart

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

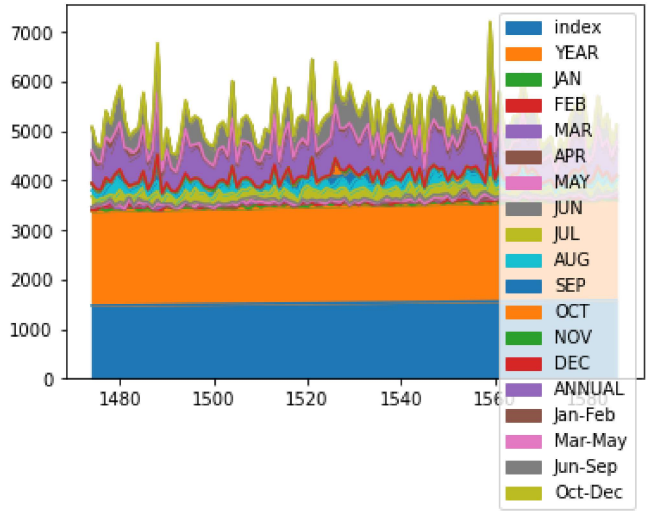
Out[10]: <AxesSubplot:>



Area chart

In [11]:
```
df.plot.area()
```

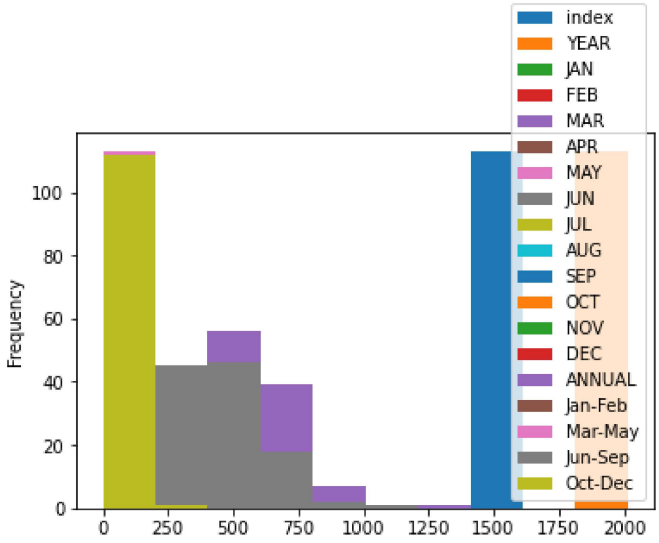
Out[11]: <AxesSubplot:>



Histogram

In [12]:
```
df.plot.hist()
```

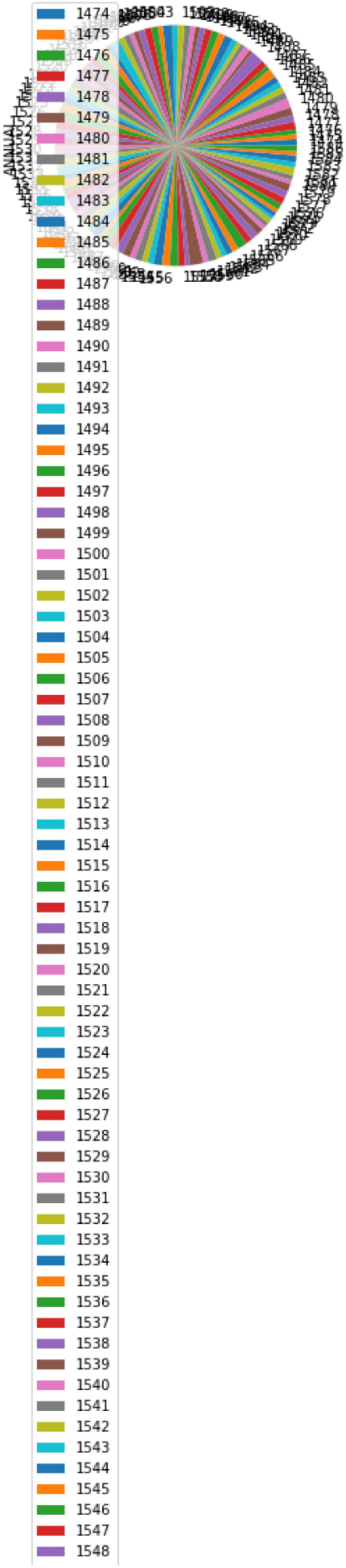
Out[12]:
```
<AxesSubplot:ylabel='Frequency'>
```



pie chart

In [13]:
```
df.plot.pie(y="ANNUAL")
```

Out[13]:
```
<AxesSubplot:ylabel='ANNUAL'>
```

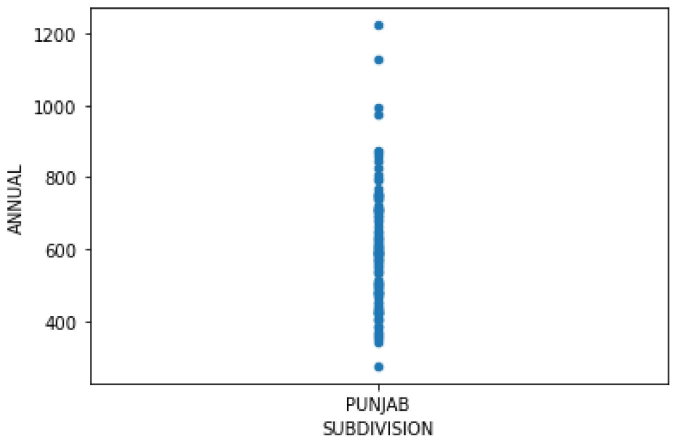

1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548

1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586

Scatter chart

In [15]:
```
df.plot.scatter(y='ANNUAL',x='SUBDIVISION')
```

Out[15]:
```
<AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>
```



In [16]:
```
df.describe()
```

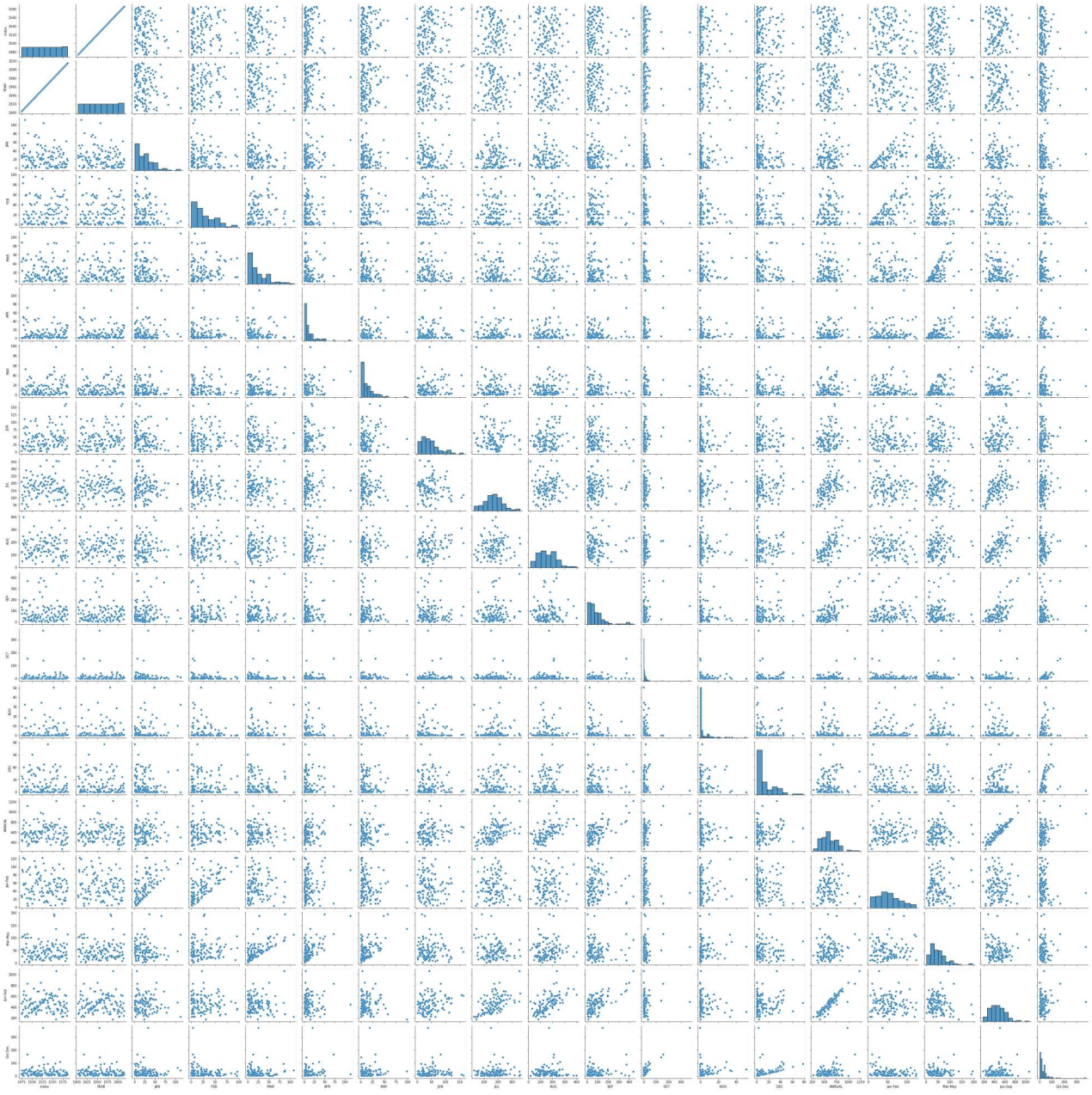
Out[16]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	
count	113.00000	113.00000	113.000000	113.000000	113.000000	113.000000	113.000000	113.000000	113.000000	113
mean	1530.00000	1959.00000	25.200000	26.810619	23.759292	12.769027	13.901770	46.755752	169.265487	158
std	32.76431	32.76431	22.192313	23.451376	23.056277	16.870029	15.212677	33.469826	68.202456	74
min	1474.00000	1903.00000	0.100000	0.000000	0.000000	0.000000	0.100000	1.600000	26.400000	17
25%	1502.00000	1931.00000	8.000000	5.700000	6.900000	2.700000	3.300000	21.700000	122.600000	96
50%	1530.00000	1959.00000	21.600000	21.300000	15.800000	6.700000	8.800000	40.700000	167.400000	150
75%	1558.00000	1987.00000	35.800000	40.300000	34.000000	15.900000	19.300000	60.300000	209.200000	208
max	1586.00000	2015.00000	112.100000	96.000000	108.500000	113.200000	98.300000	162.900000	359.300000	399

EDA AND VISUALIZATION

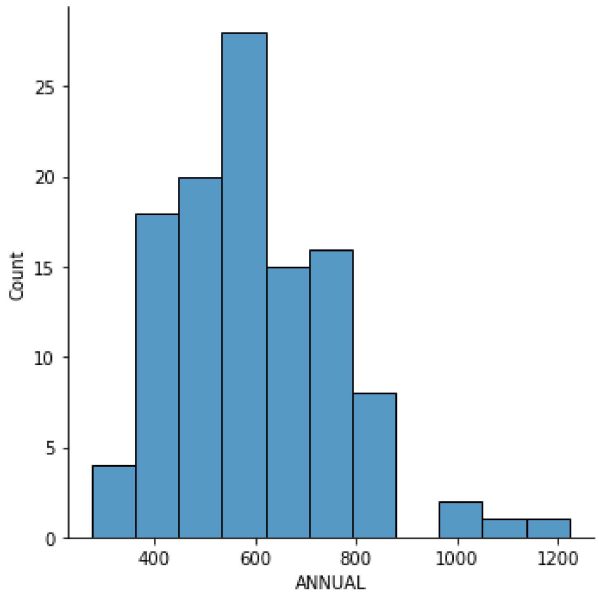
In [17]:
```
sns.pairplot(df)
```

Out[17]:
```
<seaborn.axisgrid.PairGrid at 0x1a498fc2c10>
```



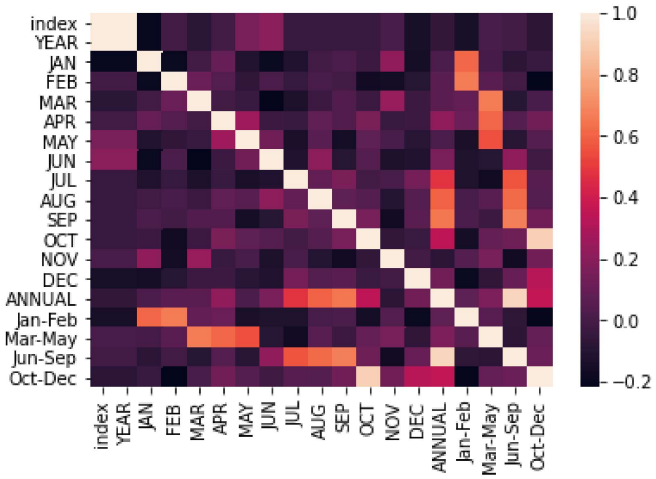
In [18]:
```
sns.displot(df['ANNUAL'])
```

Out[18]: <seaborn.axisgrid.FacetGrid at 0x1a4a749efd0>



In [19]:
```
sns.heatmap(df.corr())
```

Out[19]: <AxesSubplot:>



In [ ]: