Importing Libraries

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

Importing Datasets

In [3]:
```
df = pd.read_csv(r"C:\Users\user\Downloads\New folder\Andaman_Nicobar.csv")
df
```

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	C
0	0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	38
1	1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2	19
2	2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0	18
3	3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4	22
4	4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0	26
...	...	...	...	...	...	...	...	...	...	...	...	...	...
105	105	ANDAMAN & NICOBAR ISLANDS	2011	265.9	84.8	272.8	111.4	326.5	383.2	583.2	441.5	757.1	21
106	106	ANDAMAN & NICOBAR ISLANDS	2012	119.9	45.6	30.9	55.8	533.9	458.2	317.3	369.6	868.9	20
107	107	ANDAMAN & NICOBAR ISLANDS	2013	67.1	37.6	43.0	46.3	509.3	777.0	564.8	336.7	473.6	45
108	108	ANDAMAN & NICOBAR ISLANDS	2014	41.9	8.6	0.0	11.1	238.0	416.6	467.6	321.6	412.9	40
109	109	ANDAMAN & NICOBAR ISLANDS	2015	126.8	7.6	3.1	138.2	331.9	346.4	328.9	480.0	523.3	25

110 rows × 20 columns

Data Cleaning and Data Preprocessing

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

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')
```

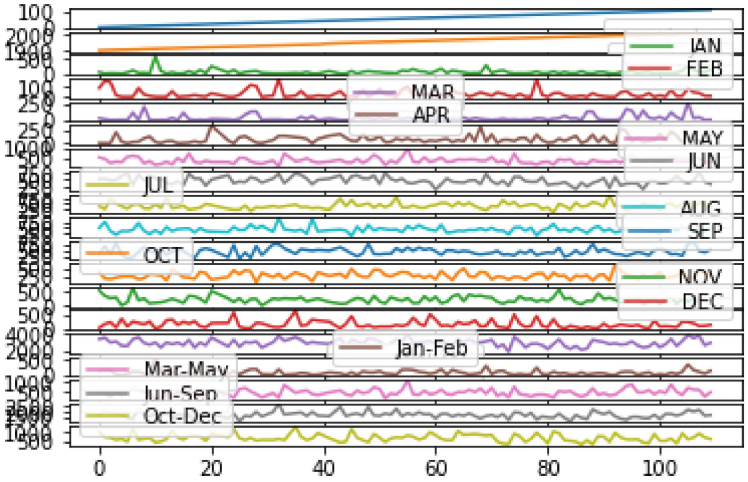
In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 104 entries, 0 to 109
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   index        104 non-null    int64  
 1   SUBDIVISION  104 non-null    object 
 2   YEAR         104 non-null    int64  
 3   JAN          104 non-null    float64
 4   FEB          104 non-null    float64
 5   MAR          104 non-null    float64
 6   APR          104 non-null    float64
 7   MAY          104 non-null    float64
 8   JUN          104 non-null    float64
 9   JUL          104 non-null    float64
 10  AUG          104 non-null    float64
 11  SEP          104 non-null    float64
 12  OCT          104 non-null    float64
 13  NOV          104 non-null    float64
 14  DEC          104 non-null    float64
 15  ANNUAL       104 non-null    float64
 16  Jan-Feb      104 non-null    float64
 17  Mar-May      104 non-null    float64
 18  Jun-Sep      104 non-null    float64
 19  Oct-Dec      104 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 17.1+ KB
```

Line chart

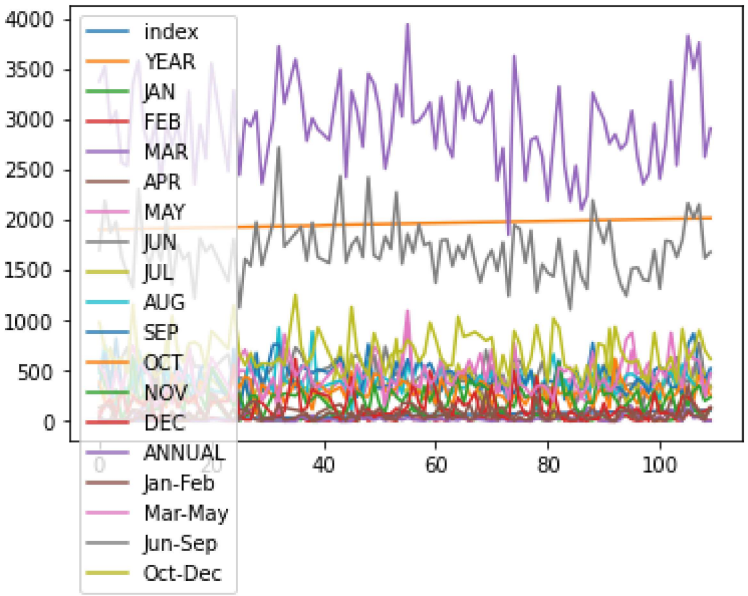
In [7]:
```
df.plot.line(subplots=True)
```

Out[7]:
```
array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>], dtype=object)
```



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

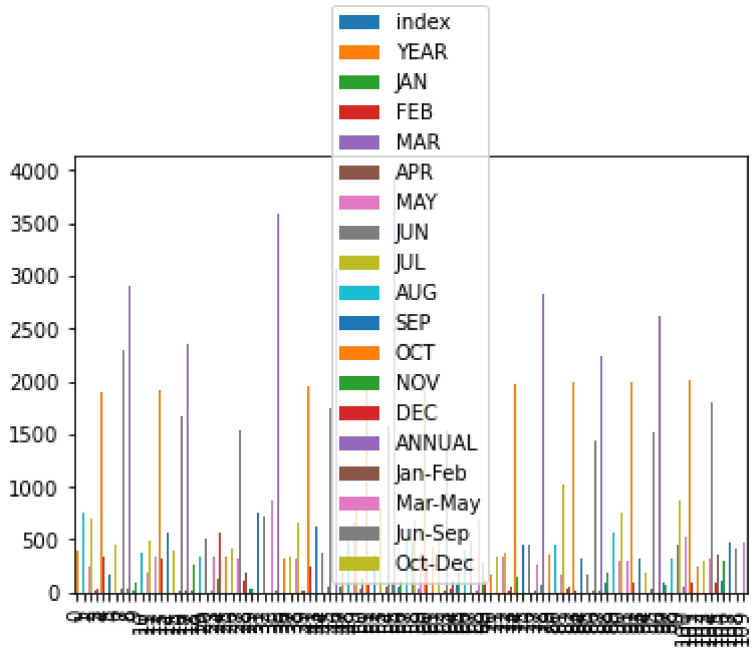
Out[8]:
```
<AxesSubplot:>
```



Bar chart

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

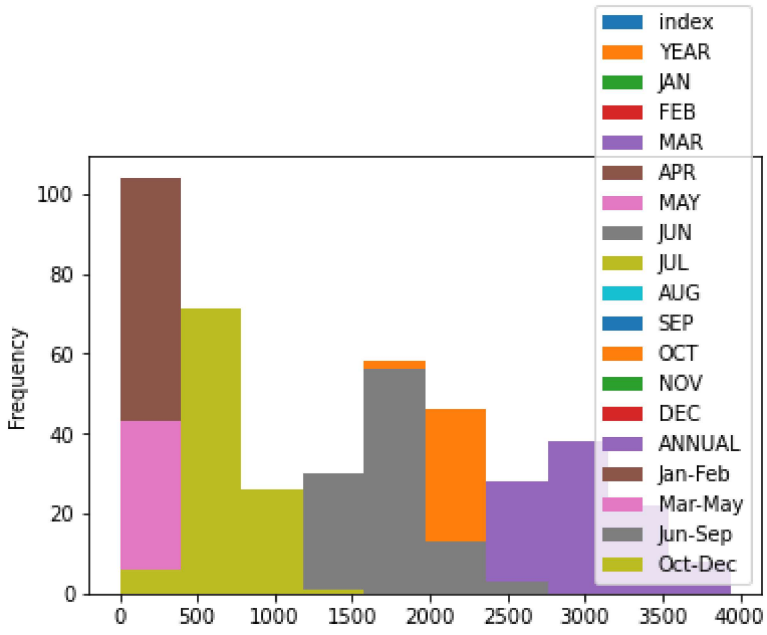
Out[9]: <AxesSubplot:>



Histogram

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

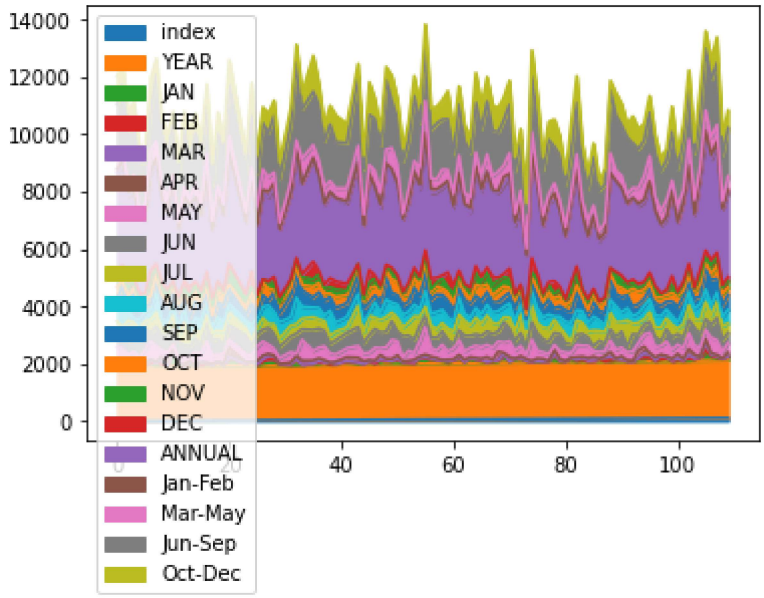
Out[10]: <AxesSubplot:ylabel='Frequency'>



Area chart

In [22]:
```
df.plot.area()
```

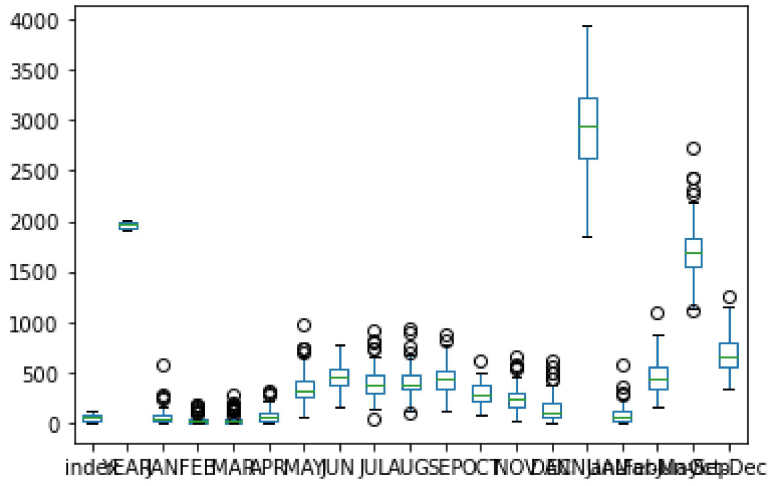
Out[22]: <AxesSubplot:>



Box plot

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

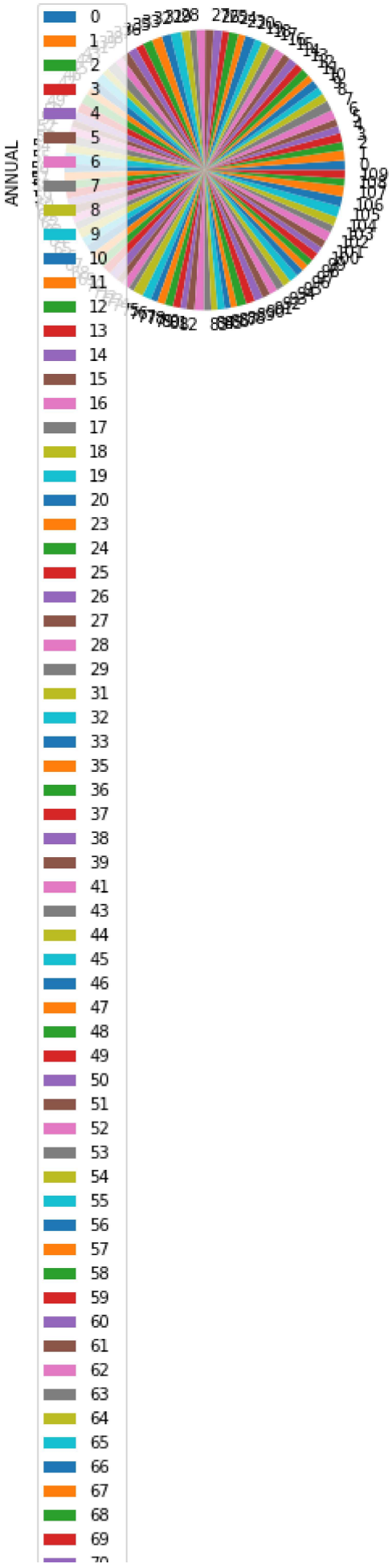
Out[12]: <AxesSubplot:>

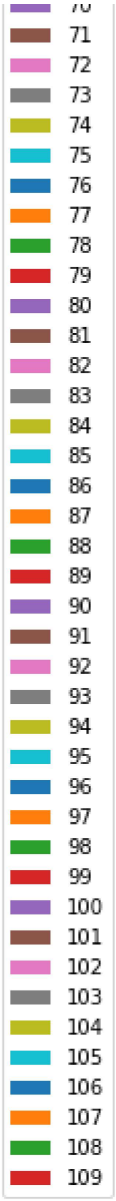


pie chart

In [15]:
```
df.plot.pie(y='ANNUAL')
```

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

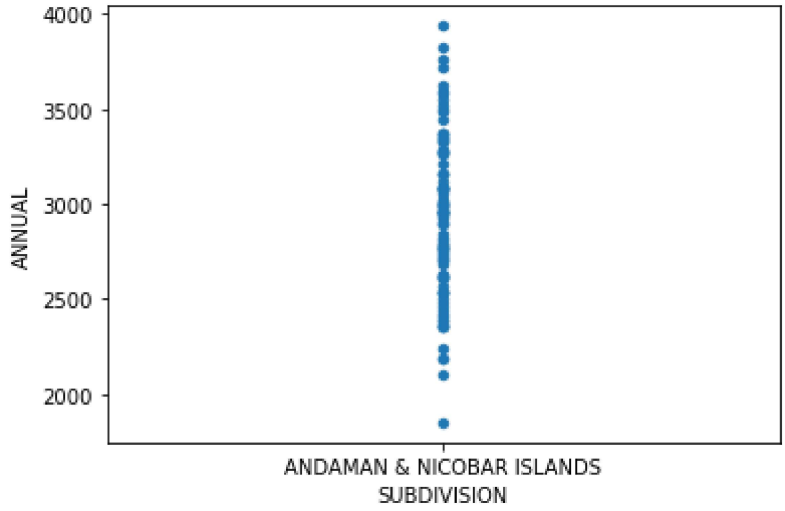

ANNUAL
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
23
24
25
26
27
28
29
31
32
33
35
36
37
38
39
41
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69



Scatter chart

In [16]:
```
df.plot.scatter(x='SUBDIVISION',y='ANNUAL')
```

Out[16]:
```
<AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>
```



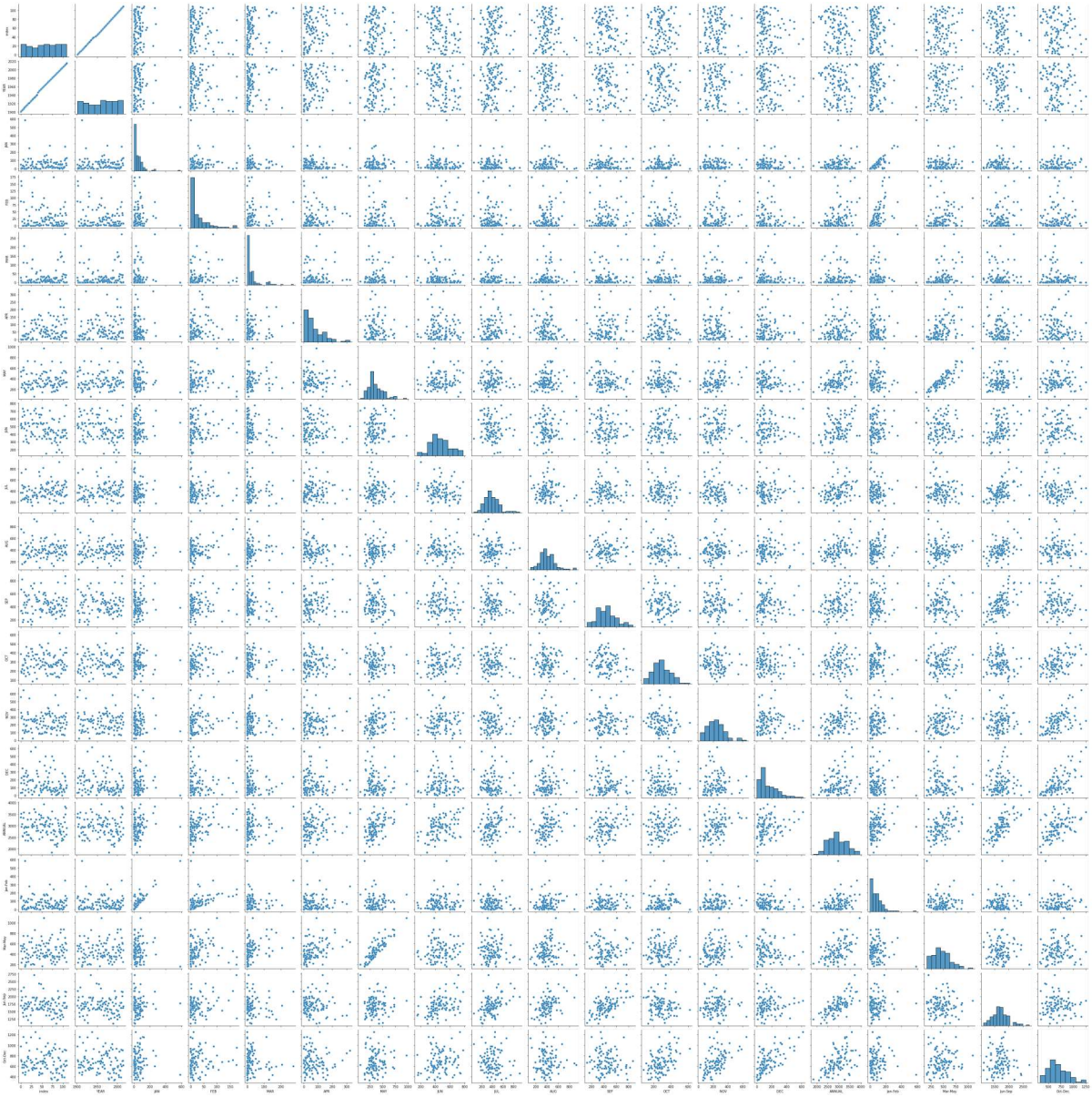
In [17]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 104 entries, 0 to 109
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   index        104 non-null    int64  
 1   SUBDIVISION  104 non-null    object 
 2   YEAR         104 non-null    int64  
 3   JAN          104 non-null    float64
 4   FEB          104 non-null    float64
 5   MAR          104 non-null    float64
 6   APR          104 non-null    float64
 7   MAY          104 non-null    float64
 8   JUN          104 non-null    float64
 9   JUL          104 non-null    float64
 10  AUG          104 non-null    float64
 11  SEP          104 non-null    float64
 12  OCT          104 non-null    float64
 13  NOV          104 non-null    float64
 14  DEC          104 non-null    float64
 15  ANNUAL       104 non-null    float64
 16  Jan-Feb      104 non-null    float64
 17  Mar-May      104 non-null    float64
 18  Jun-Sep      104 non-null    float64
 19  Oct-Dec      104 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 17.1+ KB
```

EDA AND VISUALIZATION

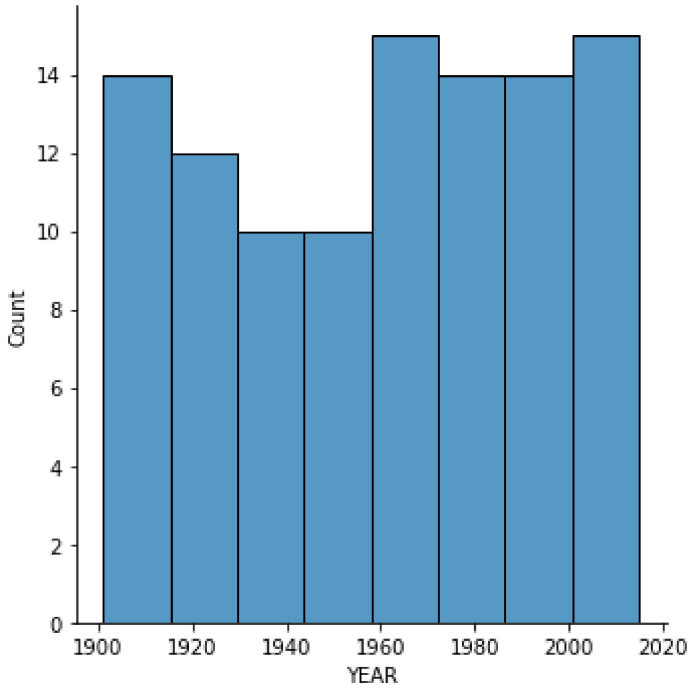
In [18]:
```
sns.pairplot(df)
```

Out[18]:
```
<seaborn.axisgrid.PairGrid at 0x1f5337bfe50>
```



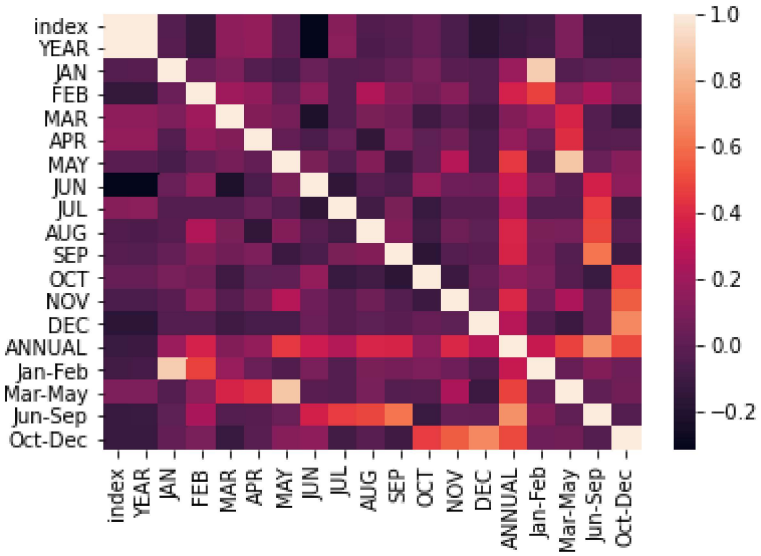
In [20]:
```
sns.displot(df['YEAR'])
```

Out[20]: <seaborn.axisgrid.FacetGrid at 0x1f53ecfaee0>



In [21]:
```
sns.heatmap(df.corr())
```

Out[21]: <AxesSubplot:>



In [ ]: