Importing Libraries

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

Importing Datasets

In [2]:
```
df=pd.read_csv("west_madhya_pradesh.csv")
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NO
0	2047	WEST MADHYA PRADESH	1901	25.8	5.8	5.8	2.8	2.1	41.2	228.9	349.9	47.9	5.6	0
1	2048	WEST MADHYA PRADESH	1902	22.1	8.4	0.0	2.0	5.9	35.9	401.9	179.4	194.1	37.9	10
2	2049	WEST MADHYA PRADESH	1903	5.3	0.0	0.0	0.0	22.3	50.6	304.9	261.1	250.2	55.1	0
3	2050	WEST MADHYA PRADESH	1904	3.2	15.5	14.8	0.0	12.0	96.6	273.0	218.6	125.9	3.3	1
4	2051	WEST MADHYA PRADESH	1905	3.5	4.4	1.1	0.8	3.0	36.1	326.3	137.6	183.5	0.3	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
110	2157	WEST MADHYA PRADESH	2011	0.0	1.7	0.1	1.8	3.6	241.5	306.7	343.3	165.0	0.2	0
111	2158	WEST MADHYA PRADESH	2012	6.2	0.0	0.0	0.9	3.1	48.2	439.2	341.2	194.3	2.1	0
112	2159	WEST MADHYA PRADESH	2013	1.7	31.1	8.5	2.8	0.4	263.7	485.1	432.6	98.9	68.7	0
113	2160	WEST MADHYA PRADESH	2014	25.6	34.4	4.6	1.4	1.4	30.6	337.4	211.0	192.6	7.0	3
114	2161	WEST MADHYA PRADESH	2015	40.2	6.4	53.5	13.3	2.0	154.1	428.2	276.6	55.6	11.0	0

115 rows × 20 columns

Data Cleaning and Data Preprocessing

In [3]:
```
df=df.dropna()
df
```

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NO
0	2047	WEST MADHYA PRADESH	1901	25.8	5.8	5.8	2.8	2.1	41.2	228.9	349.9	47.9	5.6	0
1	2048	WEST MADHYA PRADESH	1902	22.1	8.4	0.0	2.0	5.9	35.9	401.9	179.4	194.1	37.9	10
2	2049	WEST MADHYA PRADESH	1903	5.3	0.0	0.0	0.0	22.3	50.6	304.9	261.1	250.2	55.1	0
3	2050	WEST MADHYA PRADESH	1904	3.2	15.5	14.8	0.0	12.0	96.6	273.0	218.6	125.9	3.3	1
4	2051	WEST MADHYA PRADESH	1905	3.5	4.4	1.1	0.8	3.0	36.1	326.3	137.6	183.5	0.3	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
110	2157	WEST MADHYA PRADESH	2011	0.0	1.7	0.1	1.8	3.6	241.5	306.7	343.3	165.0	0.2	0
111	2158	WEST MADHYA PRADESH	2012	6.2	0.0	0.0	0.9	3.1	48.2	439.2	341.2	194.3	2.1	0
112	2159	WEST MADHYA PRADESH	2013	1.7	31.1	8.5	2.8	0.4	263.7	485.1	432.6	98.9	68.7	0
113	2160	WEST MADHYA PRADESH	2014	25.6	34.4	4.6	1.4	1.4	30.6	337.4	211.0	192.6	7.0	3
114	2161	WEST MADHYA PRADESH	2015	40.2	6.4	53.5	13.3	2.0	154.1	428.2	276.6	55.6	11.0	0

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')
```

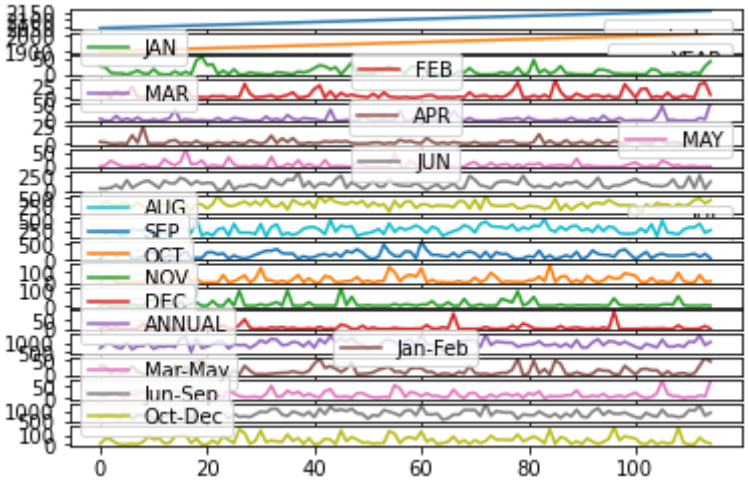
In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 114 entries, 0 to 114
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: 18.7+ KB
```

Line chart

In [6]:
```
df.plot.line(subplots=True)
```

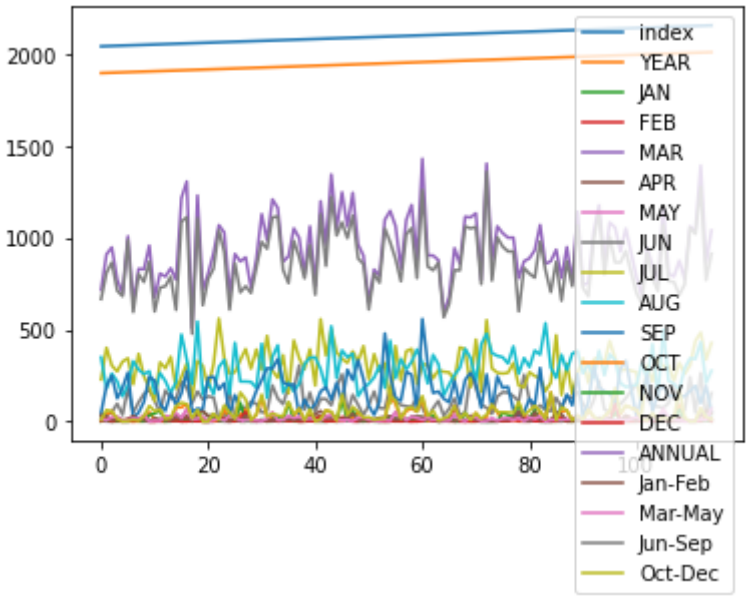
Out[6]:
```
array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>], dtype=object)
```



Line chart

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

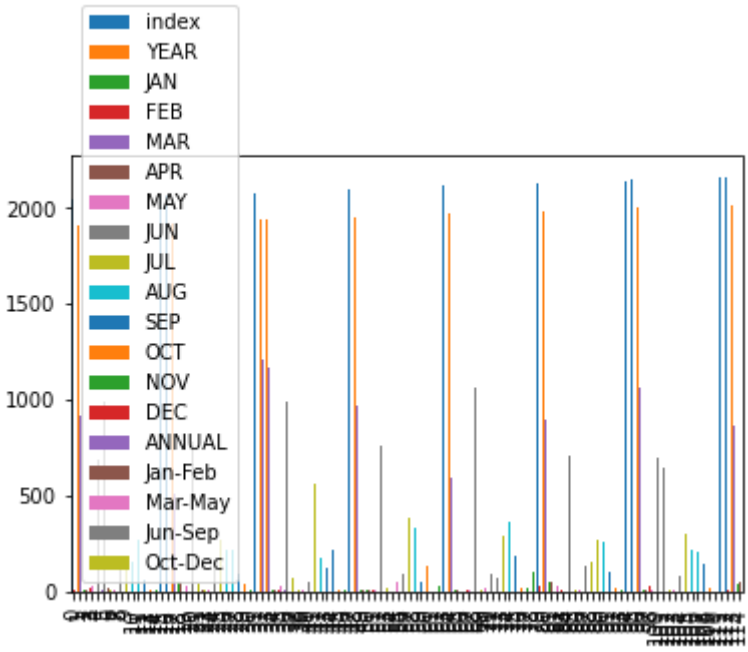
Out[7]:
```
<AxesSubplot:>
```



Bar chart

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

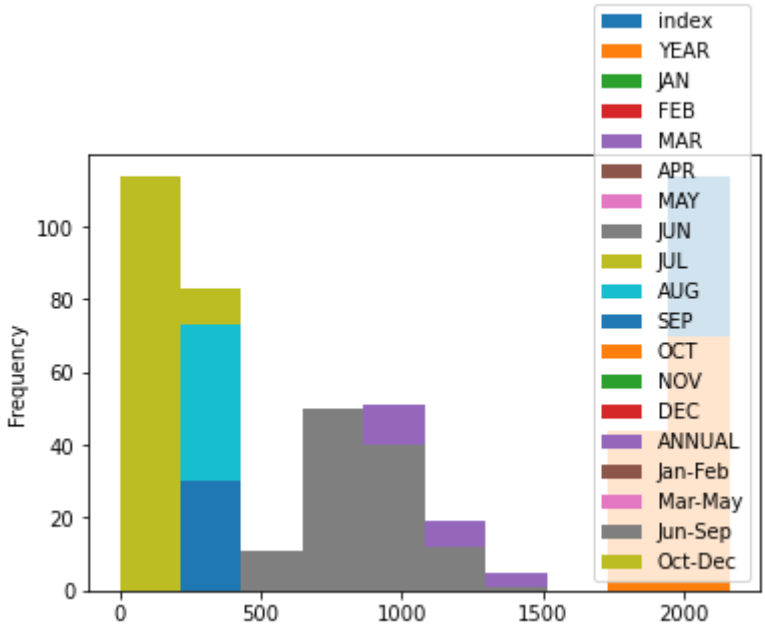
Out[8]: <AxesSubplot:>



Histogram

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

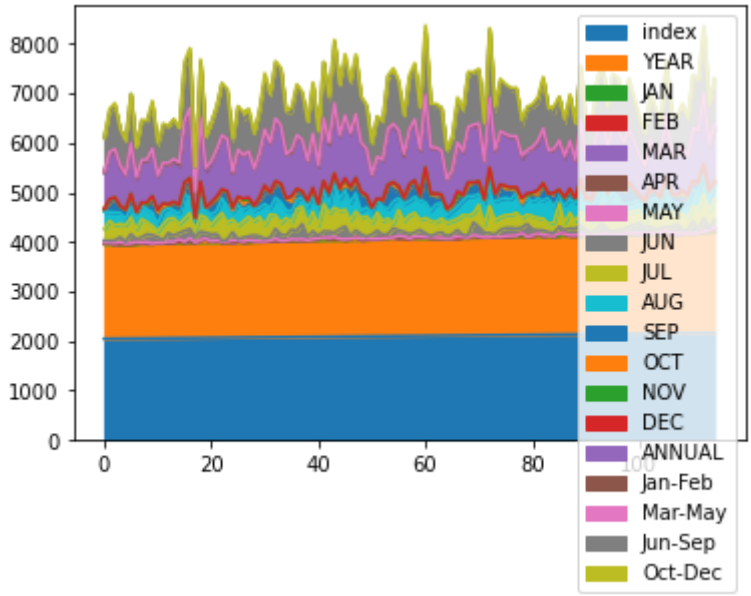
Out[9]: <AxesSubplot:ylabel='Frequency'>



Area chart

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

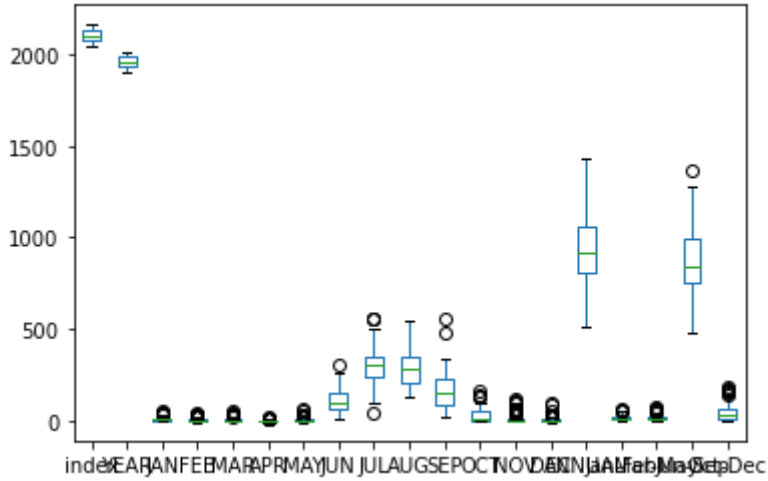
Out[10]: <AxesSubplot:>



Box chart

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

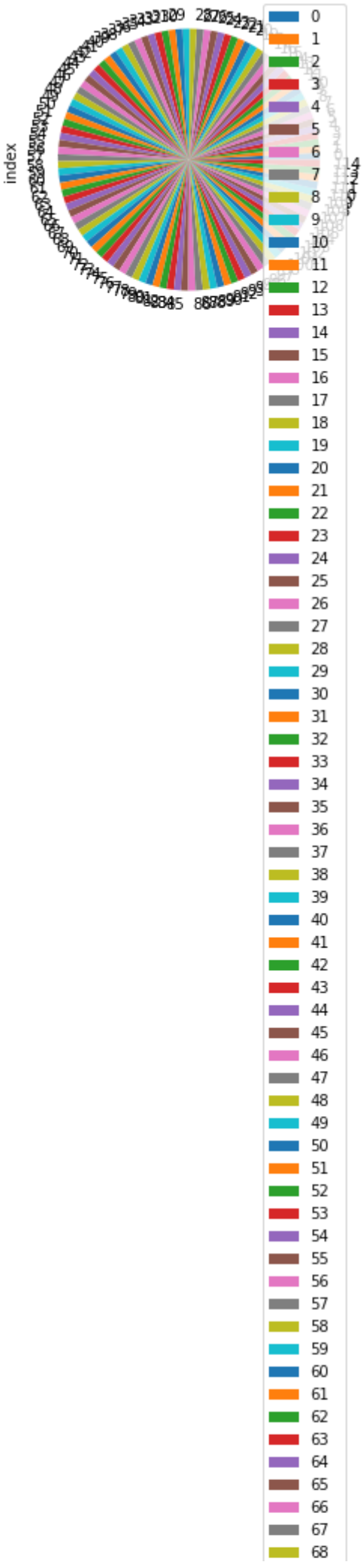
Out[11]: <AxesSubplot:>

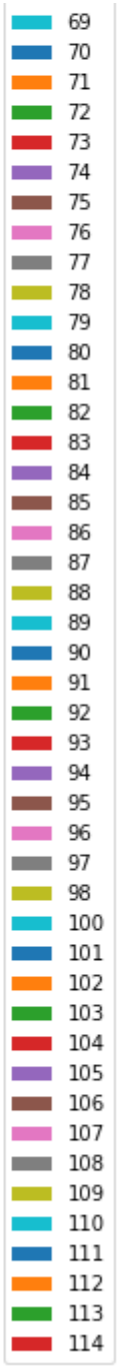


Pie chart

In [12]:
```
df.plot.pie(y='index')
```

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

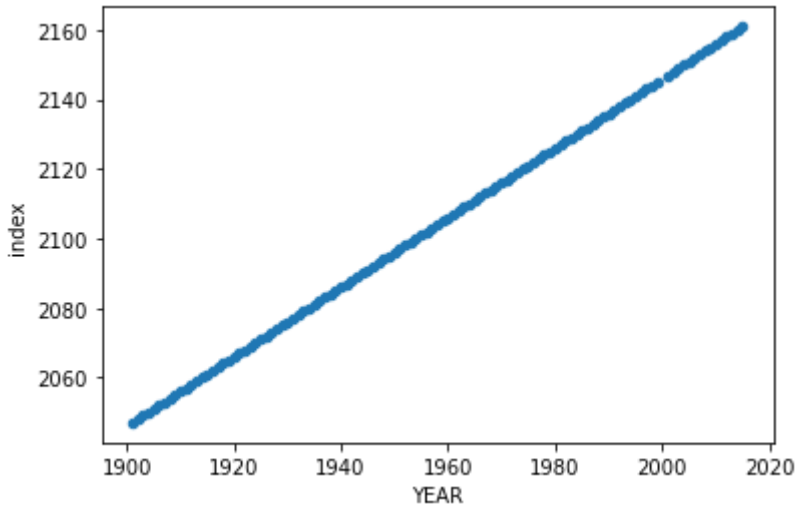

index
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68



Scatter chart

In [13]:
```
df.plot.scatter(x='YEAR' ,y='index')
```

Out[13]:
```
<AxesSubplot:xlabel='YEAR', ylabel='index'>
```



In [14]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 114 entries, 0 to 114
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: 18.7+ KB
```

In [15]:
```
df.describe()
```

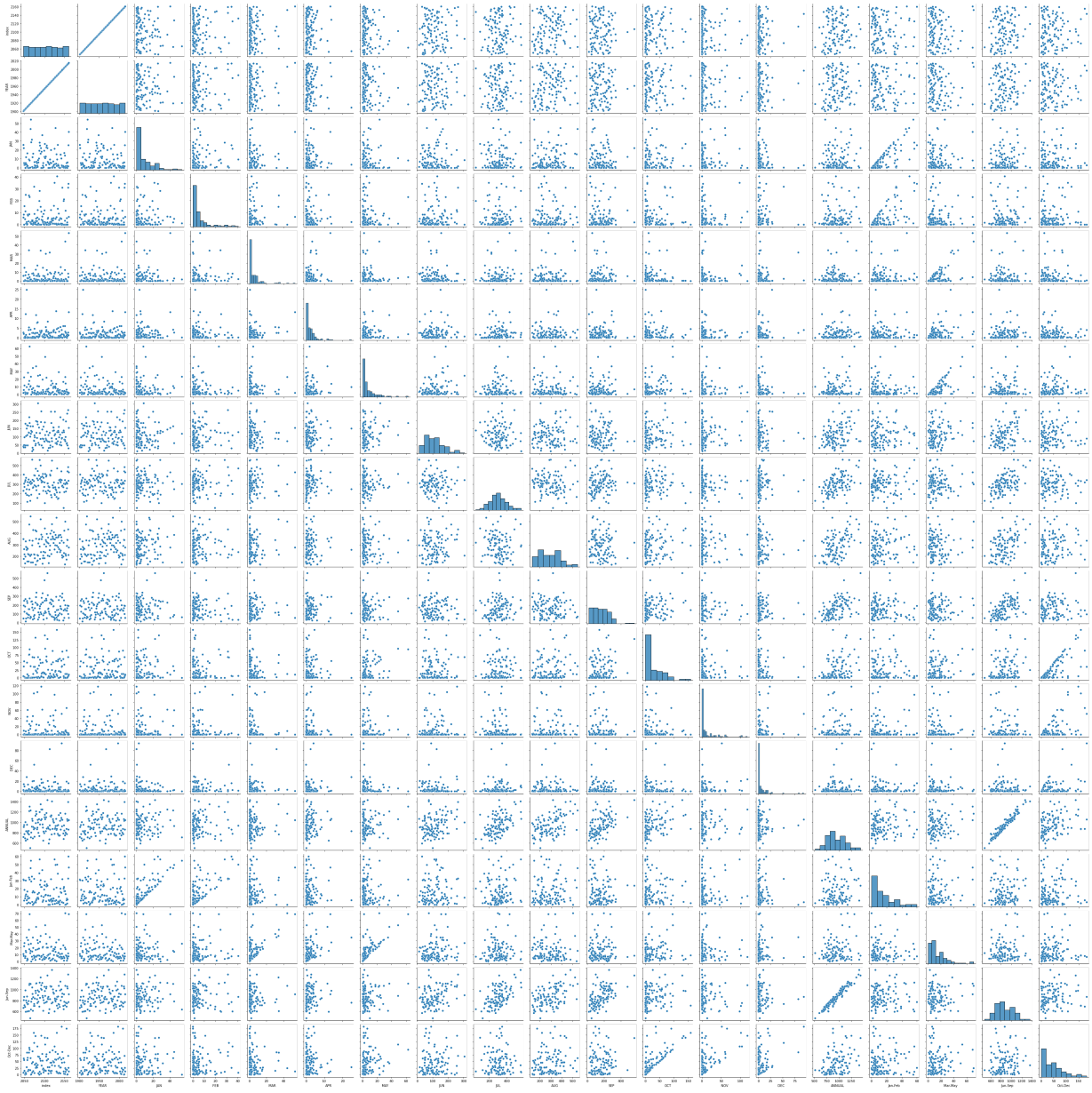
Out[15]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	
count	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.00(
mean	2103.631579	1957.631579	9.321930	6.307895	5.217544	2.395614	7.460526	111.94
std	33.252923	33.252923	11.274584	8.993755	8.973109	3.491922	10.230153	61.06
min	2047.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	12.10(
25%	2075.250000	1929.250000	0.925000	0.525000	0.225000	0.200000	1.325000	64.87
50%	2103.500000	1957.500000	5.000000	2.800000	2.050000	1.400000	3.500000	100.20(
75%	2131.750000	1985.750000	14.700000	8.200000	6.400000	3.000000	9.675000	148.75(
max	2161.000000	2015.000000	54.100000	40.500000	53.500000	24.800000	62.700000	306.30(

EDA AND VISUALIZATION

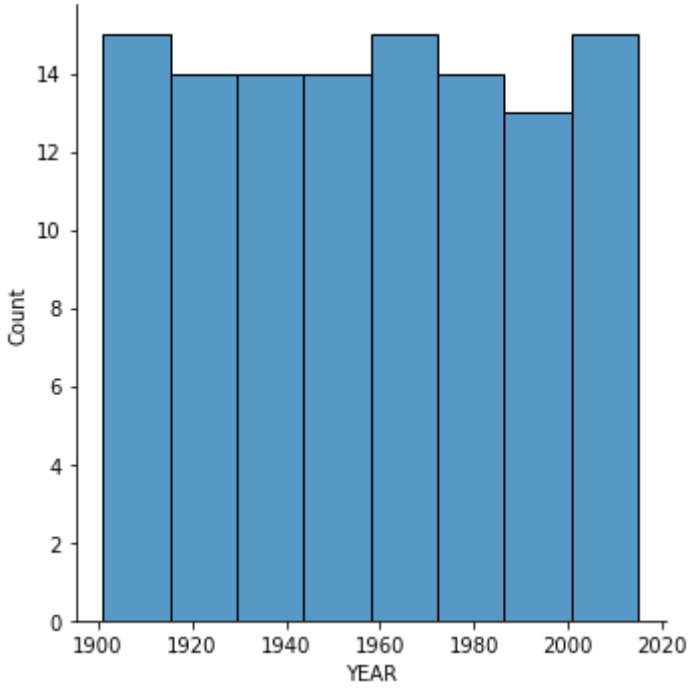
In [16]:
```
sns.pairplot(df)
```

Out[16]:
```
<seaborn.axisgrid.PairGrid at 0x215dbae5a30>
```



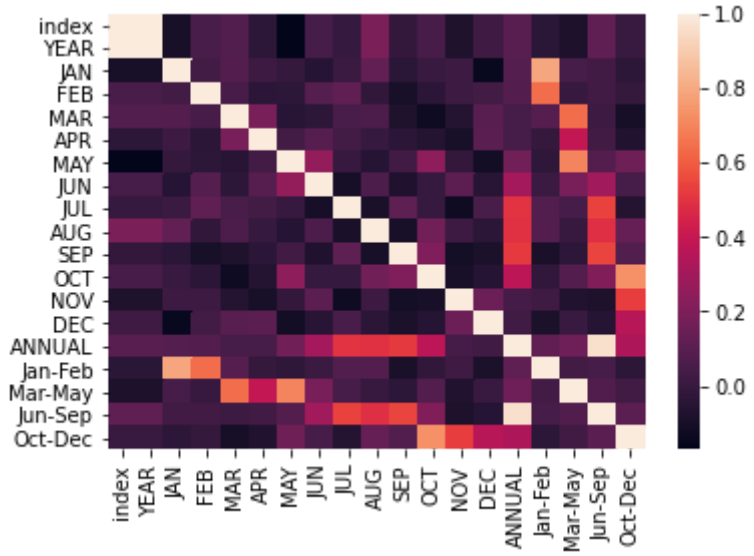
In [17]:
```
sns.displot(df['YEAR'])
```

Out[17]:
```
<seaborn.axisgrid.FacetGrid at 0x215e6ee8eb0>
```



In [18]:
```
sns.heatmap(df.corr())
```

Out[18]: <AxesSubplot:>



In [ ]: