Importing Libraries

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

Importing Datasets

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

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	N(
0	2737	MATATHWADA	1901	15.8	3.3	32.1	48.5	26.5	193.1	184.1	249.8	74.0	81.6	
1	2738	MATATHWADA	1902	1.3	0.0	0.4	7.2	0.8	52.4	120.9	85.2	273.3	61.3	8
2	2739	MATATHWADA	1903	2.6	0.8	0.0	1.7	58.3	104.4	264.2	281.9	173.3	139.9	
3	2740	MATATHWADA	1904	0.0	0.9	12.1	0.3	7.2	79.2	118.4	57.3	339.0	76.2	
4	2741	MATATHWADA	1905	1.3	2.0	0.0	6.6	4.8	84.6	94.8	137.6	157.8	15.4	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
110	2847	MATATHWADA	2011	0.0	3.8	0.7	3.5	3.1	79.2	230.1	228.5	90.0	24.8	
111	2848	MATATHWADA	2012	0.0	0.0	0.0	0.6	2.3	72.2	161.1	101.4	120.0	68.8	
112	2849	MATATHWADA	2013	1.5	9.4	2.6	7.9	6.4	160.9	293.4	136.9	154.1	94.3	
113	2850	MATATHWADA	2014	1.4	13.4	79.0	11.9	7.0	30.4	105.0	178.9	84.5	14.2	1
114	2851	MATATHWADA	2015	10.1	1.6	32.0	39.6	12.3	118.3	27.4	112.2	154.3	19.5	

115 rows × 20 columns

Data Cleaning and Data Preprocessing

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

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	N(
0	2737	MATATHWADA	1901	15.8	3.3	32.1	48.5	26.5	193.1	184.1	249.8	74.0	81.6	
1	2738	MATATHWADA	1902	1.3	0.0	0.4	7.2	0.8	52.4	120.9	85.2	273.3	61.3	8
2	2739	MATATHWADA	1903	2.6	0.8	0.0	1.7	58.3	104.4	264.2	281.9	173.3	139.9	
3	2740	MATATHWADA	1904	0.0	0.9	12.1	0.3	7.2	79.2	118.4	57.3	339.0	76.2	
4	2741	MATATHWADA	1905	1.3	2.0	0.0	6.6	4.8	84.6	94.8	137.6	157.8	15.4	

index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	N
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
110	2847	MATATHWADA	2011	0.0	3.8	0.7	3.5	3.1	79.2	230.1	228.5	90.0	24.8	
111	2848	MATATHWADA	2012	0.0	0.0	0.0	0.6	2.3	72.2	161.1	101.4	120.0	68.8	
112	2849	MATATHWADA	2013	1.5	9.4	2.6	7.9	6.4	160.9	293.4	136.9	154.1	94.3	
113	2850	MATATHWADA	2014	1.4	13.4	79.0	11.9	7.0	30.4	105.0	178.9	84.5	14.2	1
114	2851	MATATHWADA	2015	10.1	1.6	32.0	39.6	12.3	118.3	27.4	112.2	154.3	19.5	

115 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')
```
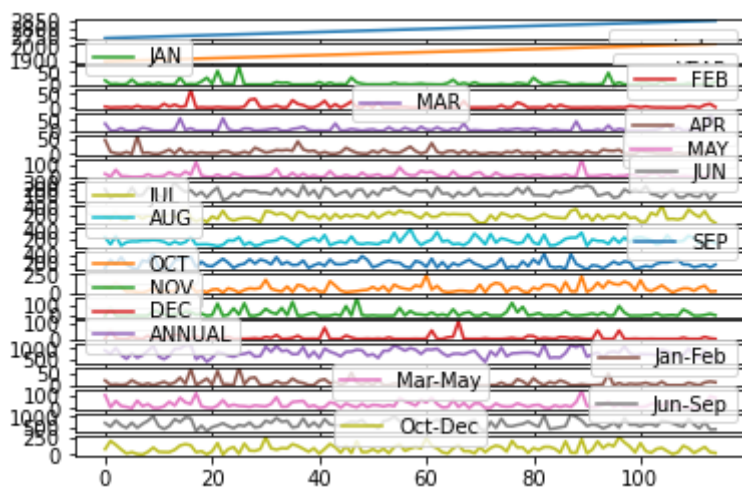
```
In [5]:   df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 115 entries, 0 to 114
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: 18.9+ 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:>,
```
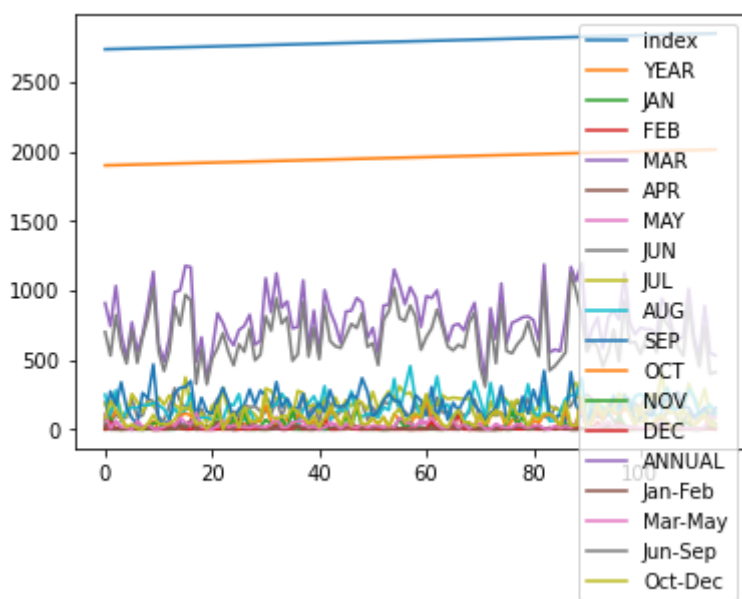
```
<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>], dtype=object)
```
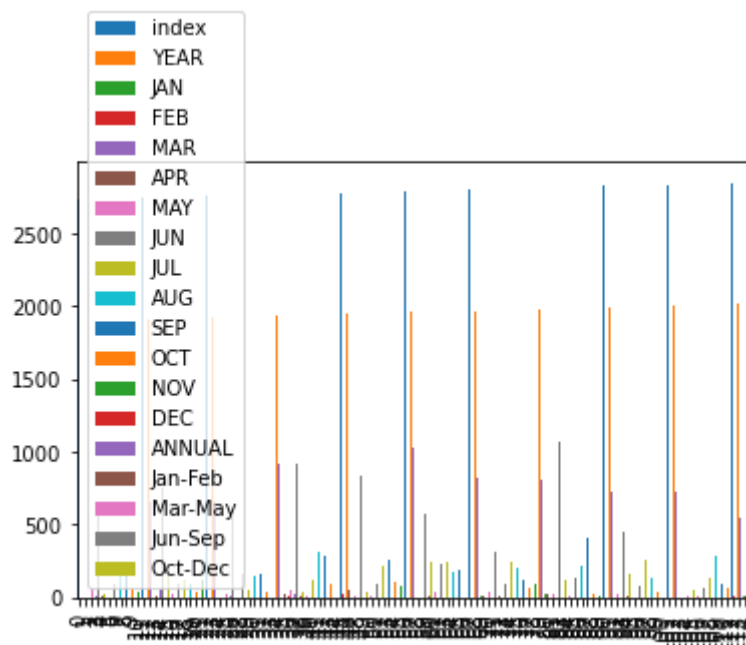

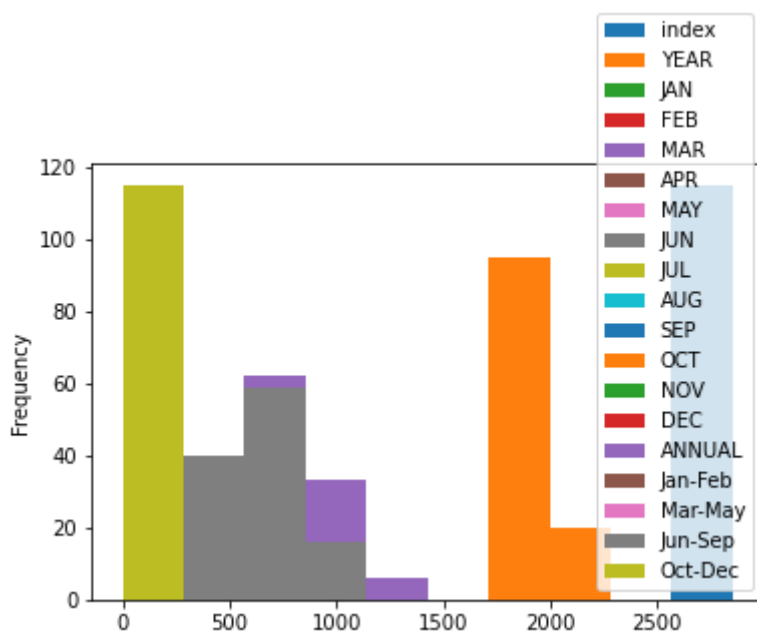# Line chart

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

Out[7]: <AxesSubplot:>

Bar chart

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

Out[8]: <AxesSubplot:>

Histogram

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

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

Area chart

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

Out[10]:   <AxesSubplot:>
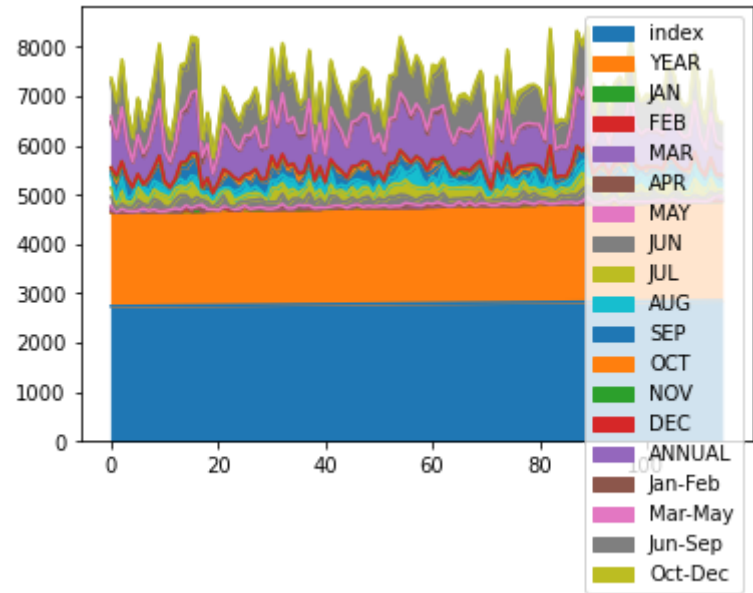
Box chart

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

Out[11]:   <AxesSubplot:>

Pie chart

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

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

▬	69
▬	70
▬	71
▬	72
▬	73
▬	74
▬	75
▬	76
▬	77
▬	78
▬	79
▬	80
▬	81
▬	82
▬	83
▬	84
▬	85
▬	86
▬	87
▬	88
▬	89
▬	90
▬	91
▬	92
▬	93
▬	94
▬	95
▬	96
▬	97
▬	98
▬	99
▬	100
▬	101
▬	102
▬	103
▬	104
▬	105
▬	106
▬	107
▬	108
▬	109
▬	110
▬	111
▬	112
▬	113
▬	114
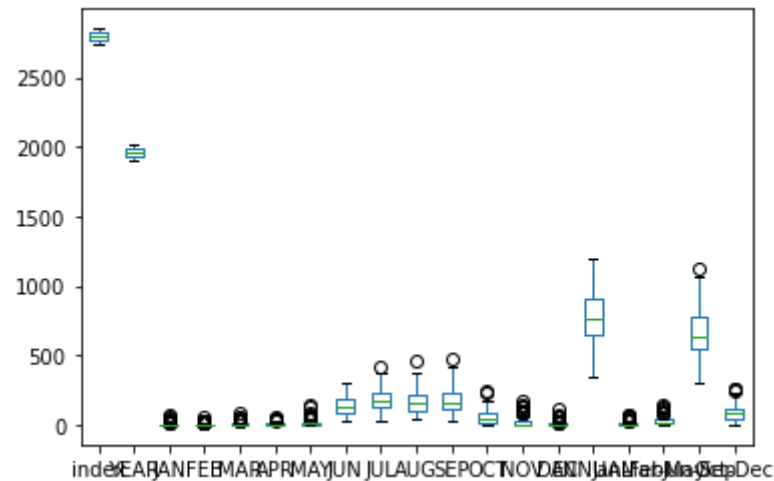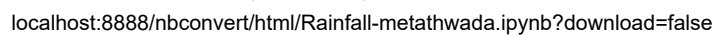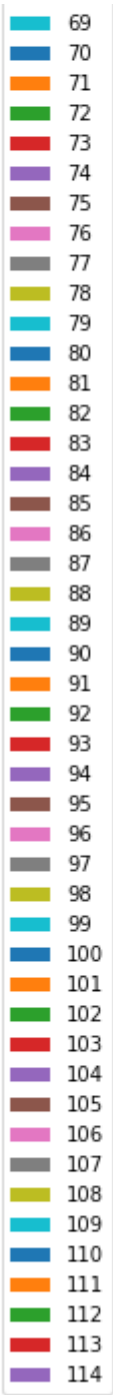
Scatter chart

In [13]:
```python
df.plot.scatter(x='YEAR' ,y='index')
```
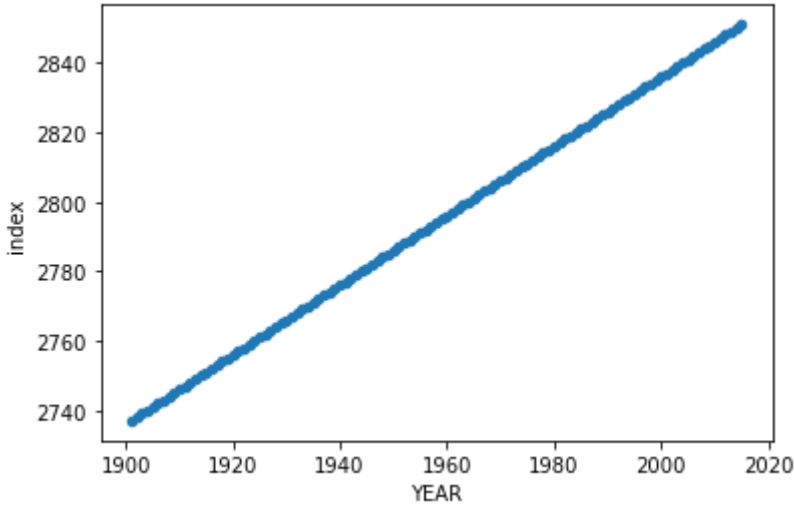
Out[13]: <AxesSubplot:xlabel='YEAR', ylabel='index'>

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

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 115 entries, 0 to 114
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: 18.9+ KB
```

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

Out[15]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	
count	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000
mean	2794.000000	1958.000000	5.000870	4.443478	7.105217	7.594783	15.646957	136.957
std	33.341666	33.341666	10.644795	8.726690	12.542271	10.028581	21.826779	57.440
min	2737.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	20.500
25%	2765.500000	1929.500000	0.000000	0.000000	0.200000	1.500000	2.200000	92.150
50%	2794.000000	1958.000000	0.900000	0.700000	2.600000	4.600000	8.000000	130.300
75%	2822.500000	1986.500000	5.600000	4.550000	8.050000	10.500000	19.200000	179.100
max	2851.000000	2015.000000	70.400000	63.500000	79.000000	61.300000	142.100000	297.000

EDA AND VISUALIZATION

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

Out[16]: <seaborn.axisgrid.PairGrid at 0x2ca73ebc1f0>
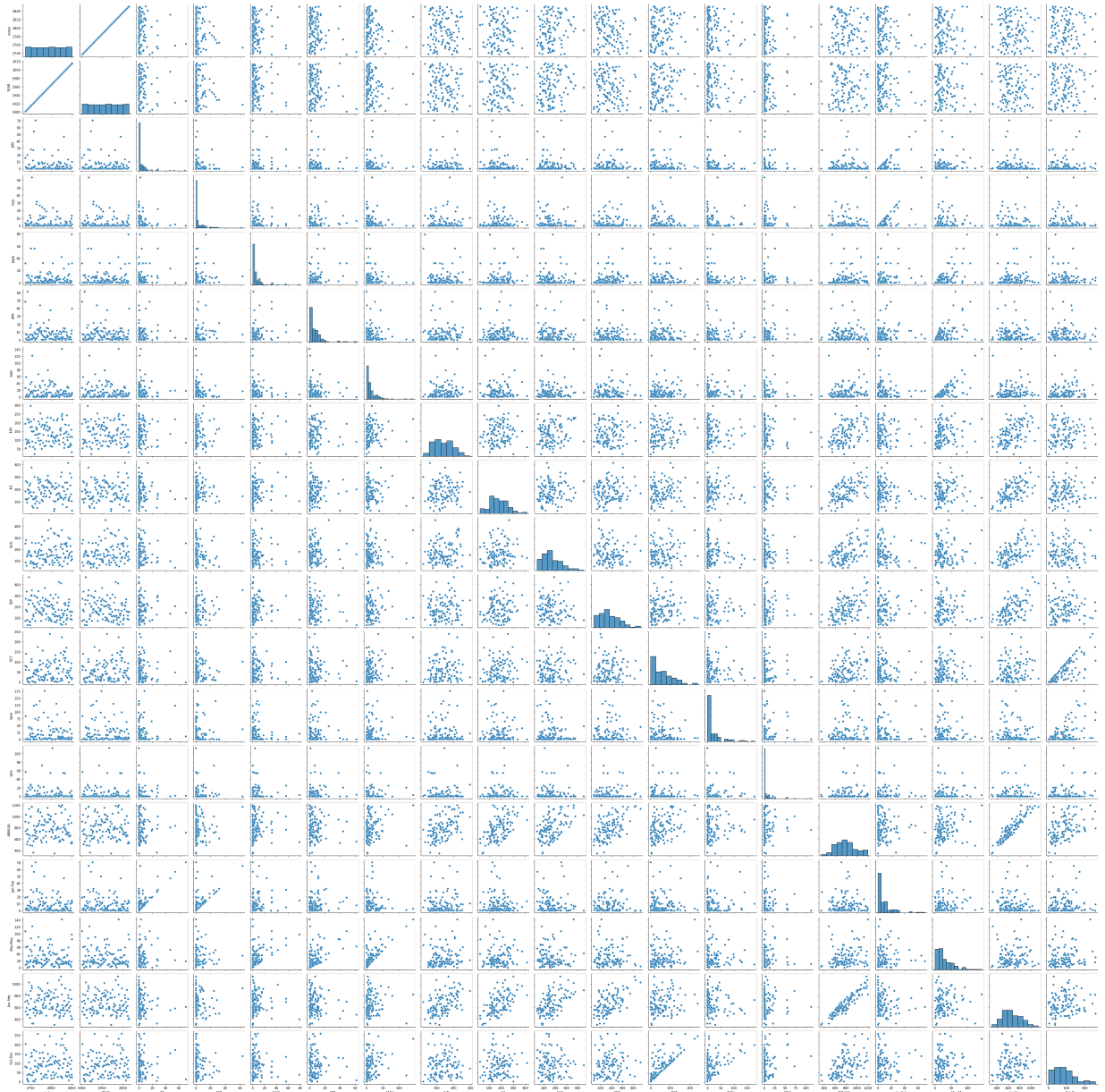
In [17]:
```python
sns.displot(df['YEAR'])
```

Out[17]: <seaborn.axisgrid.FacetGrid at 0x2ca7f43ad60>
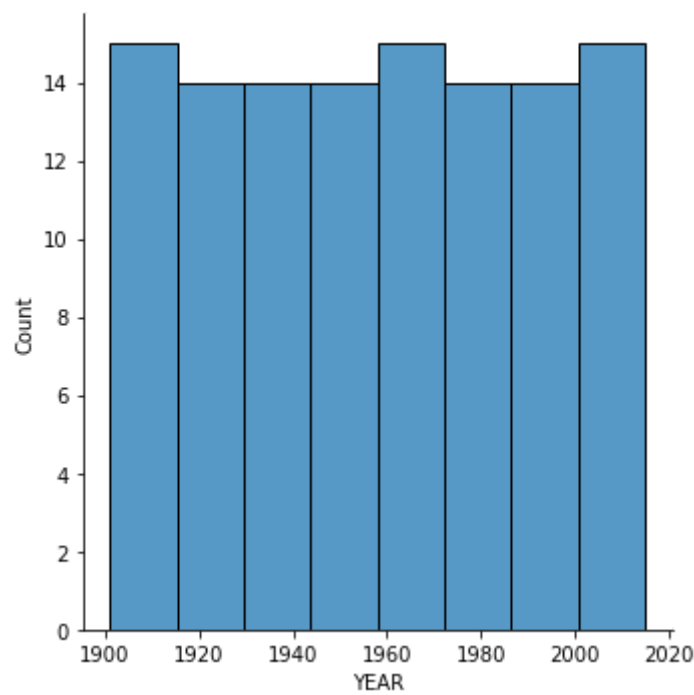
In [18]:
```python
sns.heatmap(df.corr())
```

Out[18]:   <AxesSubplot:>

In [ ]: