

# 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("uttarakhand.csv")
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
0	1242	UTTARAKHAND	1901	134.5	81.4	44.5	5.9	60.8	33.6	381.1	612.3	167.1	16.3
1	1243	UTTARAKHAND	1902	0.0	17.0	52.2	63.7	52.1	113.1	444.1	327.5	220.4	31.9
2	1244	UTTARAKHAND	1903	68.0	7.9	87.6	10.3	37.5	83.0	251.6	442.7	249.3	57.5
3	1245	UTTARAKHAND	1904	40.0	5.2	78.3	13.6	61.1	180.1	449.6	417.2	174.1	6.3
4	1246	UTTARAKHAND	1905	115.4	80.7	99.8	26.1	70.3	111.5	299.9	349.5	129.5	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
110	1352	UTTARAKHAND	2011	30.9	65.2	18.0	30.9	84.2	223.1	433.3	523.7	148.4	3.4
111	1353	UTTARAKHAND	2012	38.8	11.9	28.1	39.2	9.1	46.0	387.1	419.5	220.6	4.7
112	1354	UTTARAKHAND	2013	73.0	188.3	22.0	24.7	18.2	488.9	413.4	359.4	111.3	29.1
113	1355	UTTARAKHAND	2014	45.9	99.9	68.4	37.6	52.9	62.9	462.7	264.2	107.9	40.8
114	1356	UTTARAKHAND	2015	54.5	62.6	127.3	57.3	38.0	186.6	337.0	305.3	52.6	16.8

115 rows × 14 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
0	1242	UTTARAKHAND	1901	134.5	81.4	44.5	5.9	60.8	33.6	381.1	612.3	167.1	16.3
1	1243	UTTARAKHAND	1902	0.0	17.0	52.2	63.7	52.1	113.1	444.1	327.5	220.4	31.9
2	1244	UTTARAKHAND	1903	68.0	7.9	87.6	10.3	37.5	83.0	251.6	442.7	249.3	57.5
3	1245	UTTARAKHAND	1904	40.0	5.2	78.3	13.6	61.1	180.1	449.6	417.2	174.1	6.3
4	1246	UTTARAKHAND	1905	115.4	80.7	99.8	26.1	70.3	111.5	299.9	349.5	129.5	0.0

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
	...	...	...	...	...	...	...	...	...	...	...	...	...
110	1352	UTTARAKHAND	2011	30.9	65.2	18.0	30.9	84.2	223.1	433.3	523.7	148.4	3.4
111	1353	UTTARAKHAND	2012	38.8	11.9	28.1	39.2	9.1	46.0	387.1	419.5	220.6	4.7
112	1354	UTTARAKHAND	2013	73.0	188.3	22.0	24.7	18.2	488.9	413.4	359.4	111.3	29.1
113	1355	UTTARAKHAND	2014	45.9	99.9	68.4	37.6	52.9	62.9	462.7	264.2	107.9	40.8
114	1356	UTTARAKHAND	2015	54.5	62.6	127.3	57.3	38.0	186.6	337.0	305.3	52.6	16.8

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
```

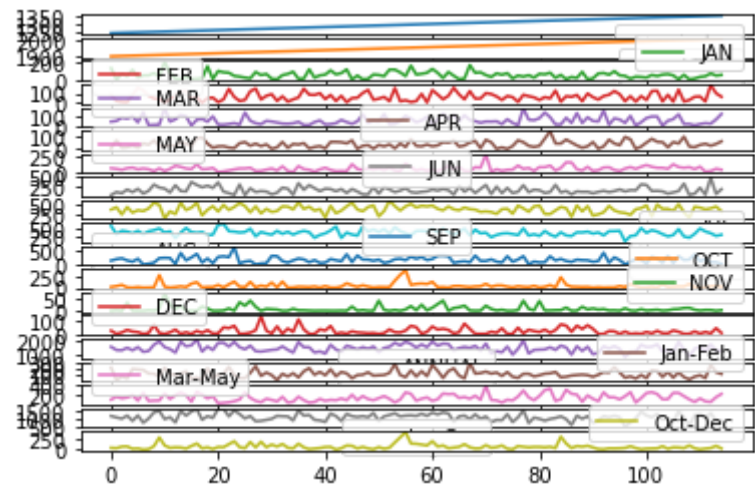
## Line chart

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:>, <AxesSubplot:>])

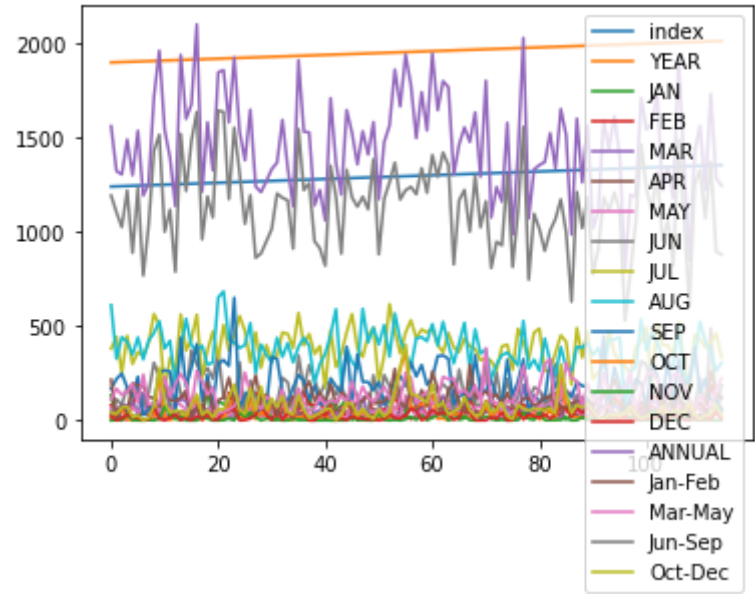
```
<AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>,
<AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>], dtype=object)
```



Line chart

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

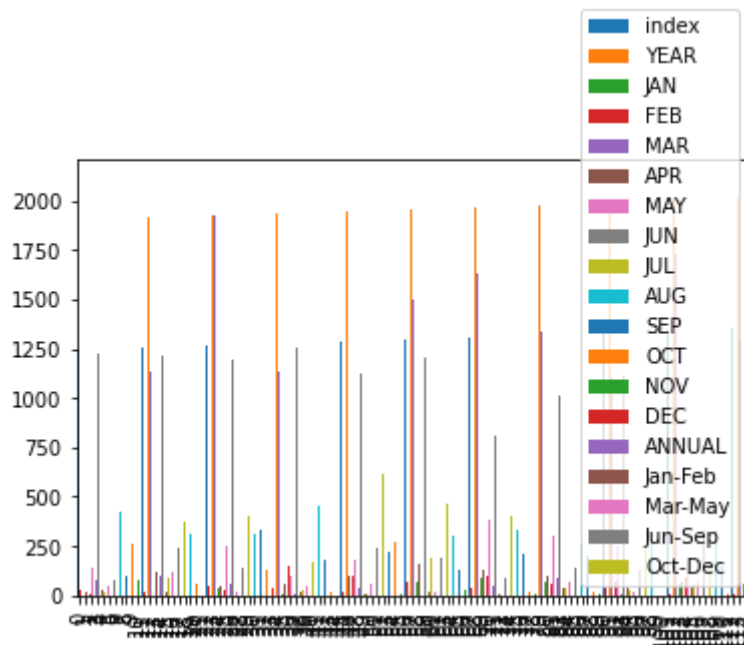
Out[7]: <AxesSubplot:~>



Bar chart

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

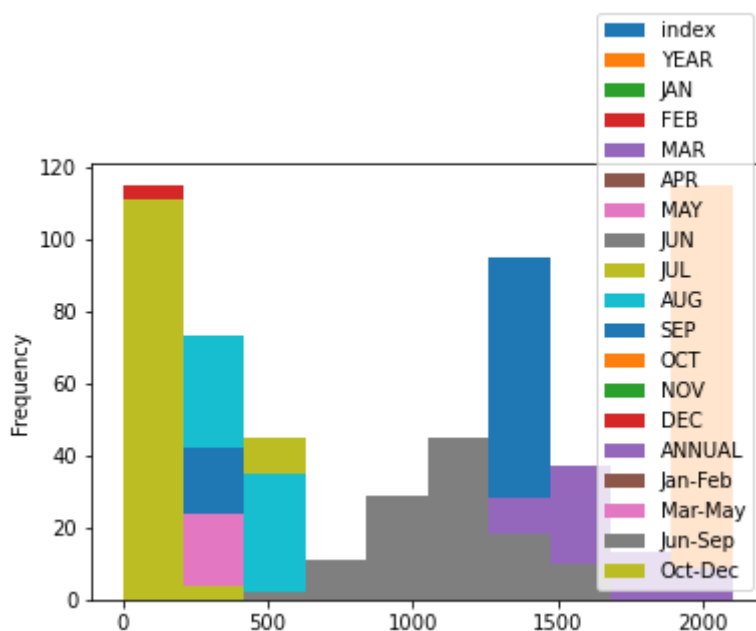
Out[8]: <AxesSubplot:~>



## Histogram

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

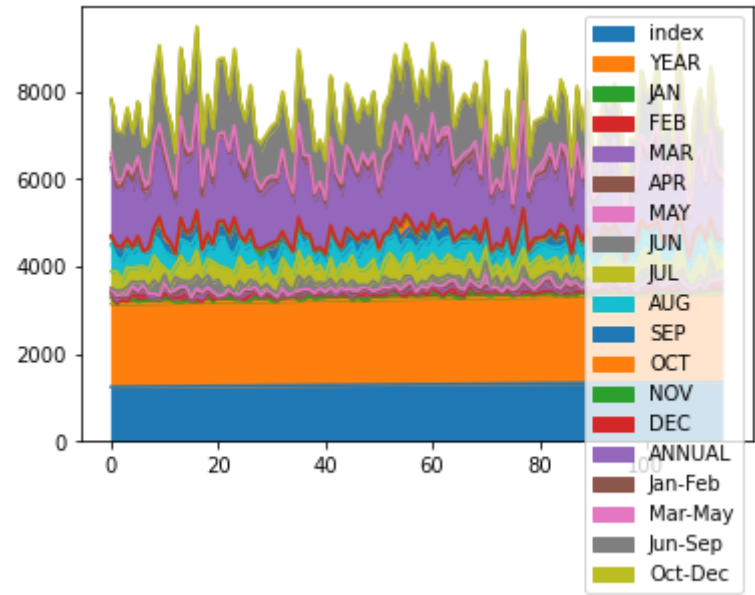
Out[9]: `<AxesSubplot:ylabel='Frequency'>`



## Area chart

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

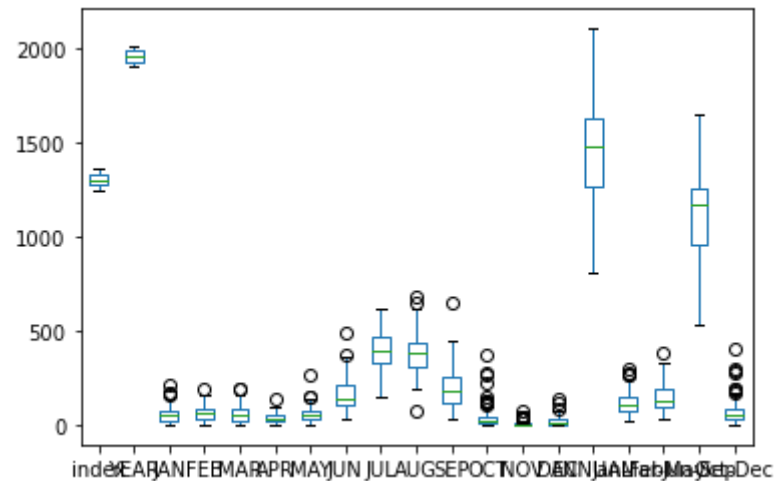
Out[10]: `<AxesSubplot:>`



## Box chart

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

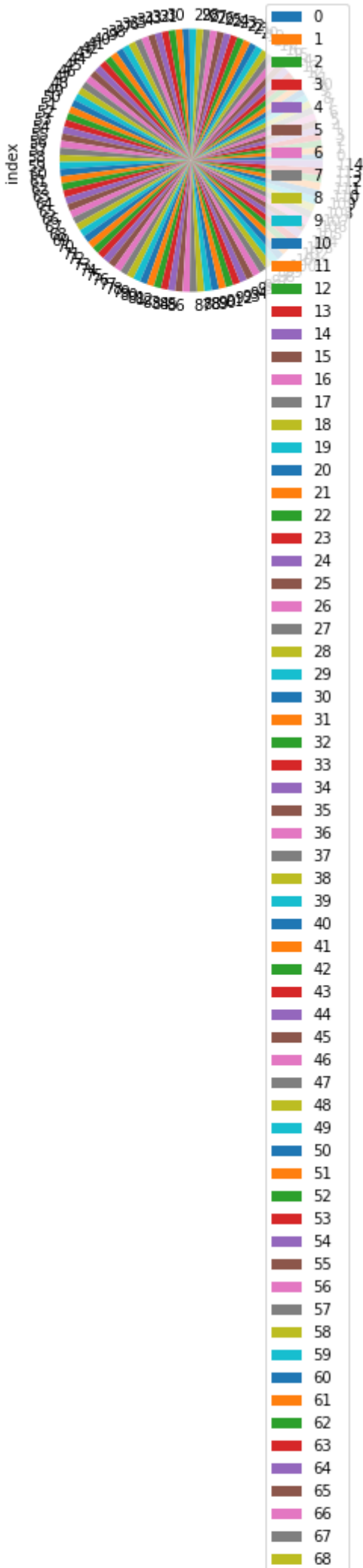
Out[11]: <AxesSubplot:>

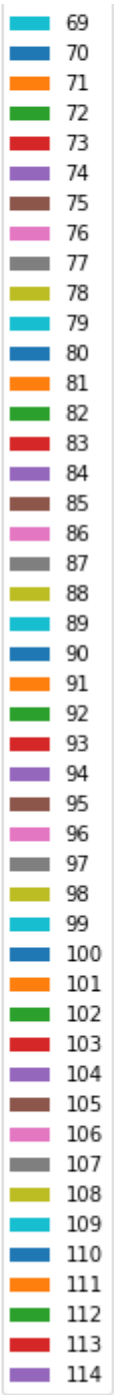


## Pie chart

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

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

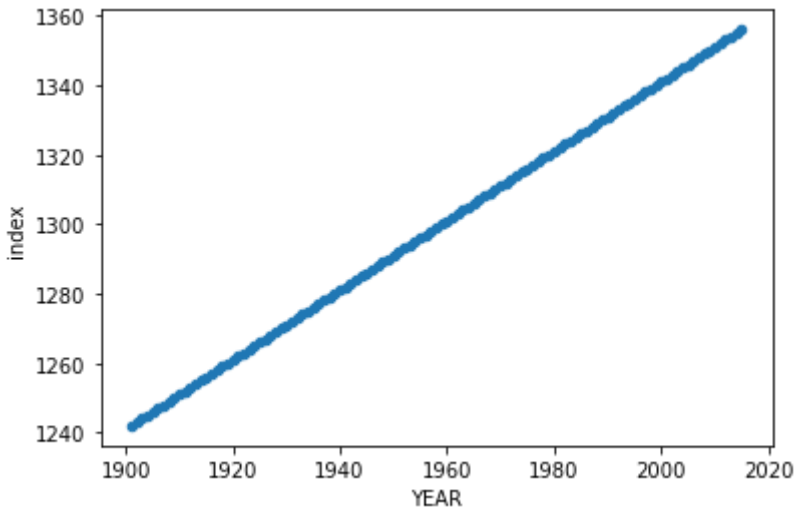




# Scatter chart

```
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: 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()
```

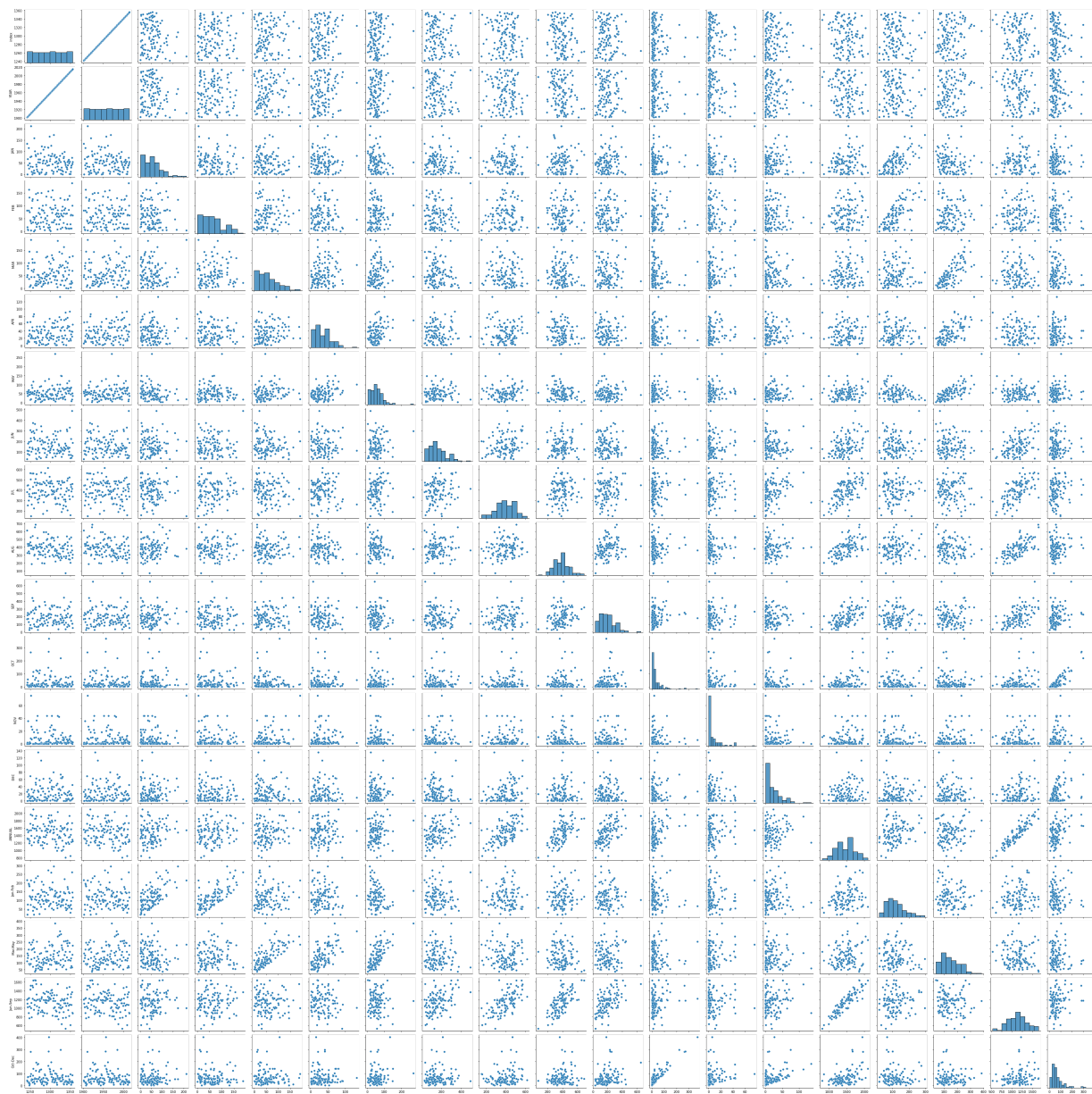
	index	YEAR	JAN	FEB	MAR	APR	MAY	
count	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000
mean	1299.000000	1958.000000	53.797391	63.452174	57.272174	35.166087	55.338261	162.551000
std	33.341666	33.341666	40.887384	44.040532	42.438752	24.116540	36.597919	86.271000
min	1242.000000	1901.000000	0.000000	0.000000	0.000000	1.100000	3.600000	33.600000
25%	1270.500000	1929.500000	21.400000	27.950000	22.850000	18.250000	28.050000	104.300000
50%	1299.000000	1958.000000	49.700000	60.100000	47.700000	30.700000	50.500000	139.900000
75%	1327.500000	1986.500000	76.200000	88.100000	80.600000	51.200000	71.450000	211.650000
max	1356.000000	2015.000000	211.400000	188.300000	190.300000	132.900000	270.200000	488.900000



# EDA AND VISUALIZATION

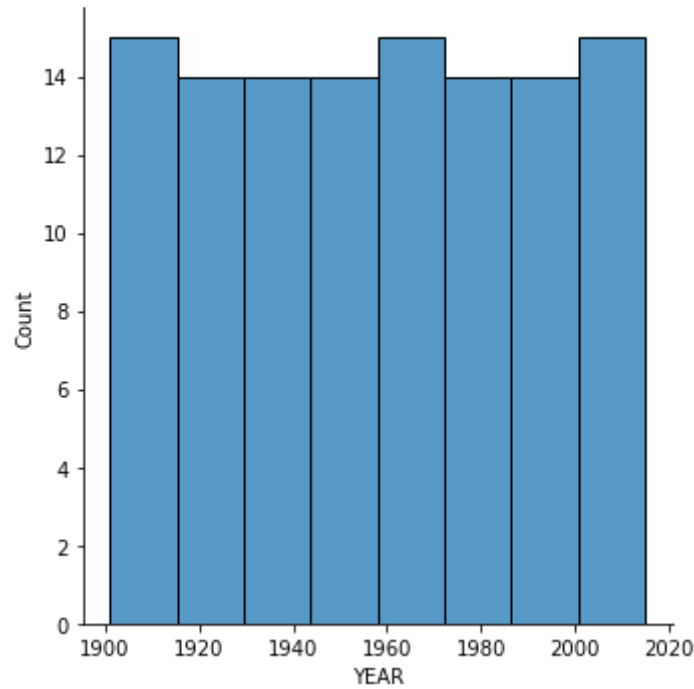
```
In [16]: sns.pairplot(df)
```

```
Out[16]: <seaborn.axisgrid.PairGrid at 0x1fe9fefae80>
```



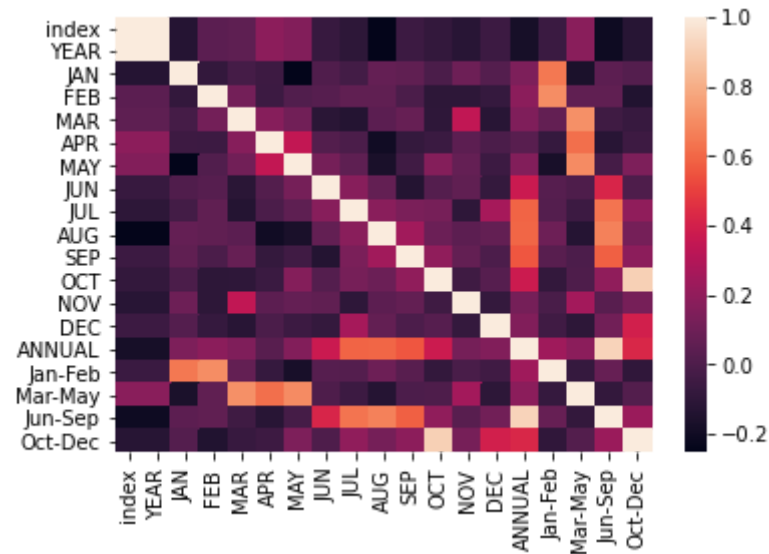
```
In [17]: sns.displot(df['YEAR'])
```

```
Out[17]: <seaborn.axisgrid.FacetGrid at 0x1fea9444e50>
```



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

Out[18]: <AxesSubplot:>



```
In [ ]:
```