

# 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("rainfall_east madhya pradesh.csv")
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV
<b>0</b>	2162	EAST MADHYA PRADESH	1901	48.5	38.1	15.7	10.7	6.2	61.0	367.5	589.2	189.9	5.9	0.0
<b>1</b>	2163	EAST MADHYA PRADESH	1902	14.9	8.9	0.0	3.6	2.7	28.0	411.9	227.0	236.6	17.0	27.6
<b>2</b>	2164	EAST MADHYA PRADESH	1903	5.6	2.9	0.3	0.9	37.5	67.5	261.4	366.7	257.4	177.9	0.0
<b>3</b>	2165	EAST MADHYA PRADESH	1904	2.0	15.3	48.2	0.0	8.6	109.9	443.2	316.6	135.6	44.8	3.2
<b>4</b>	2166	EAST MADHYA PRADESH	1905	15.9	8.0	14.3	12.3	10.2	34.4	292.4	243.3	250.9	2.9	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
<b>110</b>	2272	EAST MADHYA PRADESH	2011	0.6	1.9	0.3	7.1	4.7	332.5	323.6	326.9	276.5	1.1	0.0
<b>111</b>	2273	EAST MADHYA PRADESH	2012	39.4	0.7	0.6	1.1	1.2	67.8	398.9	351.7	172.6	12.7	3.8
<b>112</b>	2274	EAST MADHYA PRADESH	2013	2.0	43.4	14.1	9.5	0.3	311.9	456.2	480.8	78.0	124.2	0.5
<b>113</b>	2275	EAST MADHYA PRADESH	2014	32.1	49.7	17.8	5.1	2.5	91.8	283.4	231.8	139.6	56.4	1.9
<b>114</b>	2276	EAST MADHYA PRADESH	2015	37.3	11.0	73.4	25.8	6.3	139.2	262.2	272.1	71.6	38.2	1.2

115 rows × 20 columns

## head

In [3]:

```
df.head(5)
df
```

Out[3]:

		index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV
0	2162		EAST MADHYA PRADESH	1901	48.5	38.1	15.7	10.7	6.2	61.0	367.5	589.2	189.9	5.9	0.0
1	2163		EAST MADHYA PRADESH	1902	14.9	8.9	0.0	3.6	2.7	28.0	411.9	227.0	236.6	17.0	27.6
2	2164		EAST MADHYA PRADESH	1903	5.6	2.9	0.3	0.9	37.5	67.5	261.4	366.7	257.4	177.9	0.0
3	2165		EAST MADHYA PRADESH	1904	2.0	15.3	48.2	0.0	8.6	109.9	443.2	316.6	135.6	44.8	3.2
4	2166		EAST MADHYA PRADESH	1905	15.9	8.0	14.3	12.3	10.2	34.4	292.4	243.3	250.9	2.9	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
110	2272		EAST MADHYA PRADESH	2011	0.6	1.9	0.3	7.1	4.7	332.5	323.6	326.9	276.5	1.1	0.0
111	2273		EAST MADHYA PRADESH	2012	39.4	0.7	0.6	1.1	1.2	67.8	398.9	351.7	172.6	12.7	3.8
112	2274		EAST MADHYA PRADESH	2013	2.0	43.4	14.1	9.5	0.3	311.9	456.2	480.8	78.0	124.2	0.5
113	2275		EAST MADHYA PRADESH	2014	32.1	49.7	17.8	5.1	2.5	91.8	283.4	231.8	139.6	56.4	1.9
114	2276		EAST MADHYA PRADESH	2015	37.3	11.0	73.4	25.8	6.3	139.2	262.2	272.1	71.6	38.2	1.2

115 rows × 20 columns



# tail

In [4]:

```
df.tail(5)
df
```

Out[4]:

		index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	
		0	EAST MADHYA PRADESH	2162	1901	48.5	38.1	15.7	10.7	6.2	61.0	367.5	589.2	189.9	5.9	0.0
		1	EAST MADHYA PRADESH	2163	1902	14.9	8.9	0.0	3.6	2.7	28.0	411.9	227.0	236.6	17.0	27.6
		2	EAST MADHYA PRADESH	2164	1903	5.6	2.9	0.3	0.9	37.5	67.5	261.4	366.7	257.4	177.9	0.0
		3	EAST MADHYA PRADESH	2165	1904	2.0	15.3	48.2	0.0	8.6	109.9	443.2	316.6	135.6	44.8	3.2
		4	EAST MADHYA PRADESH	2166	1905	15.9	8.0	14.3	12.3	10.2	34.4	292.4	243.3	250.9	2.9	0.0
		...	...	...	...	...	...	...	...	...	...	...	...	...	...	
		110	EAST MADHYA PRADESH	2272	2011	0.6	1.9	0.3	7.1	4.7	332.5	323.6	326.9	276.5	1.1	0.0
		111	EAST MADHYA PRADESH	2273	2012	39.4	0.7	0.6	1.1	1.2	67.8	398.9	351.7	172.6	12.7	3.8
		112	EAST MADHYA PRADESH	2274	2013	2.0	43.4	14.1	9.5	0.3	311.9	456.2	480.8	78.0	124.2	0.5
		113	EAST MADHYA PRADESH	2275	2014	32.1	49.7	17.8	5.1	2.5	91.8	283.4	231.8	139.6	56.4	1.9
		114	EAST MADHYA PRADESH	2276	2015	37.3	11.0	73.4	25.8	6.3	139.2	262.2	272.1	71.6	38.2	1.2

115 rows × 20 columns



## Data Cleaning and Data Preprocessing

### describe()

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

Out[5]:

	<b>index</b>	<b>YEAR</b>	<b>JAN</b>	<b>FEB</b>	<b>MAR</b>	<b>APR</b>	<b>MAY</b>	<b>JUN</b>
<b>count</b>	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000
<b>mean</b>	2219.000000	1958.000000	19.401739	18.693913	13.637391	7.188696	9.273043	141.029565
<b>std</b>	33.341666	33.341666	22.318347	20.795522	17.354996	10.473272	12.145379	79.359922
<b>min</b>	2162.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	26.300000
<b>25%</b>	2190.500000	1929.500000	2.200000	3.650000	1.150000	1.350000	2.100000	82.850000
<b>50%</b>	2219.000000	1958.000000	12.800000	11.300000	8.000000	3.200000	5.100000	118.500000
<b>75%</b>	2247.500000	1986.500000	29.650000	27.400000	18.650000	8.750000	10.500000	197.300000
<b>max</b>	2276.000000	2015.000000	120.700000	103.100000	87.300000	72.400000	74.200000	356.600000

## shape

In [6]: `np.shape(df)`

Out[6]: (115, 20)

## size

In [7]: `np.size(df)`

Out[7]: 2300

## dropna

In [8]: `df=df.dropna()`

## columns

In [9]: `df.columns`

Out[9]: 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')

# info()

In [10]:

```
df.info()
```

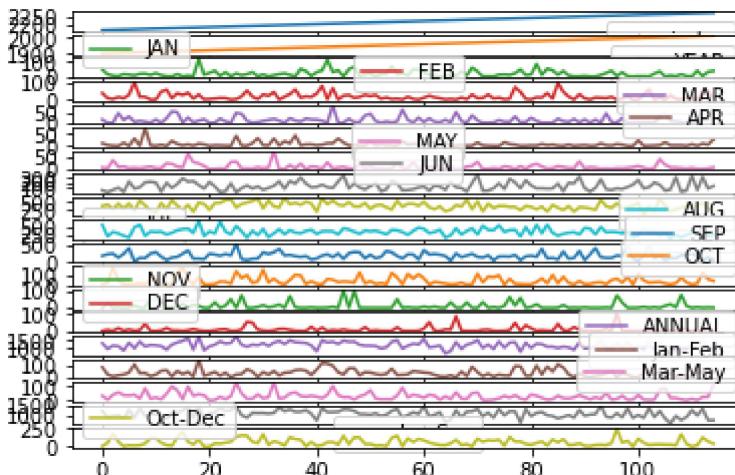
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 115 entries, 0 to 114
Data columns (total 20 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   index       115 non-null    int64  
 1   SUBDIVISION 115 non-null    object  
 2   YEAR        115 non-null    int64  
 3   JAN         115 non-null    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 [11]:

```
df.plot.line(subplots=True)
```

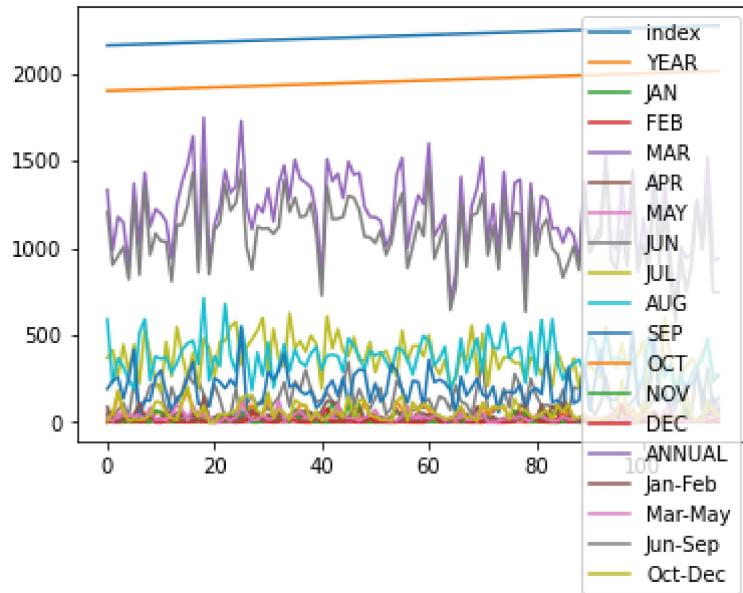
```
Out[11]: array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
   <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
   <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
   <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
   <AxesSubplot:>, <AxesSubplot:>], dtype=object)
```



# Line chart

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

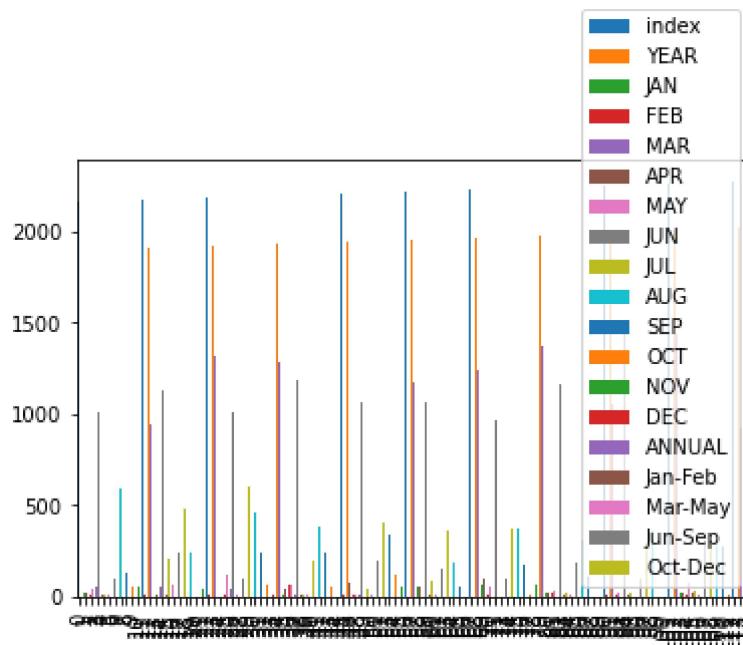
Out[12]: <AxesSubplot:>



# Bar chart

In [13]: `df.plot.bar()`

Out[13]: <AxesSubplot:>

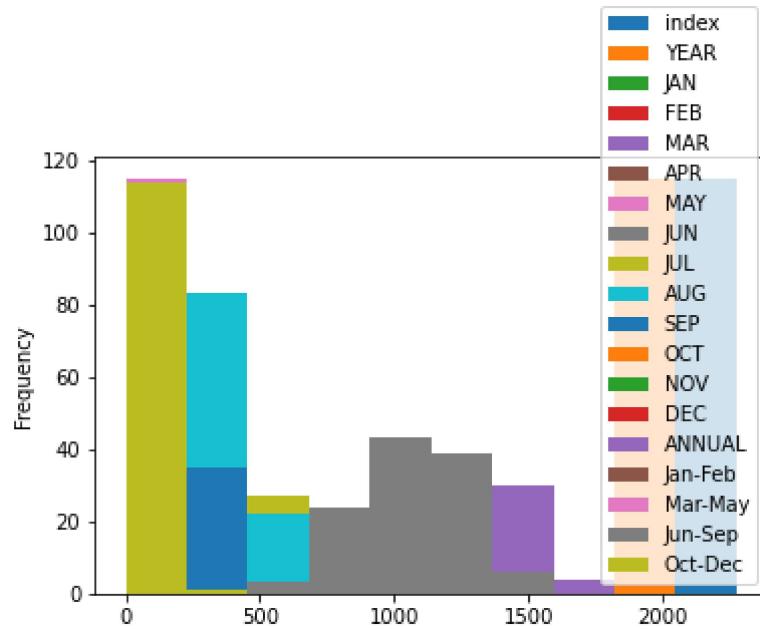


# Histogram

In [14]:

df.plot.hist()

Out[14]: &lt;AxesSubplot:ylabel='Frequency'&gt;

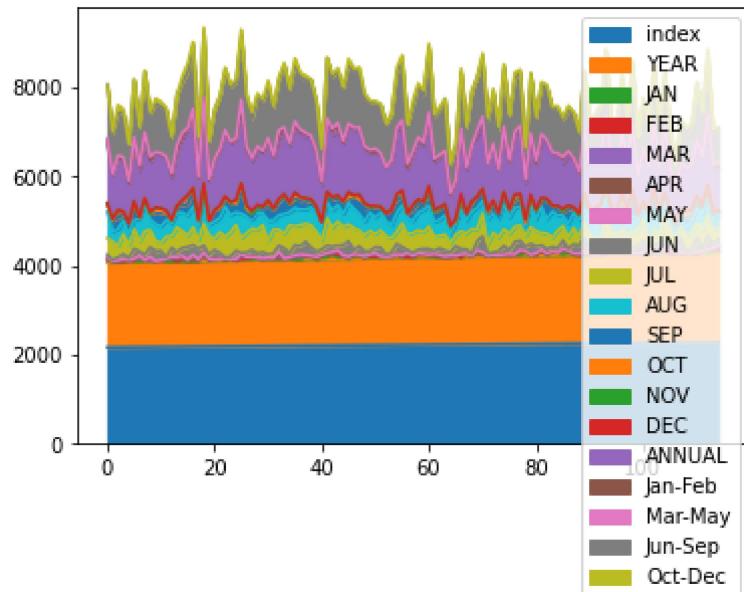


# Area chart

In [15]:

df.plot.area()

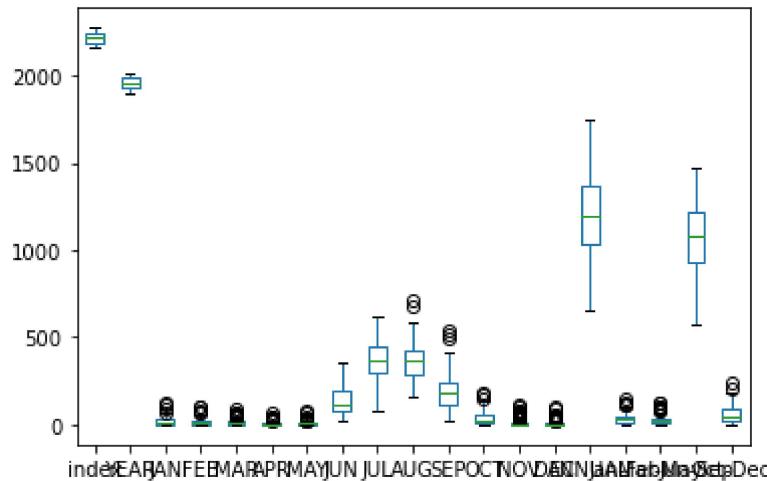
Out[15]: &lt;AxesSubplot:&gt;



# Box chart

```
In [16]: df.plot.box()
```

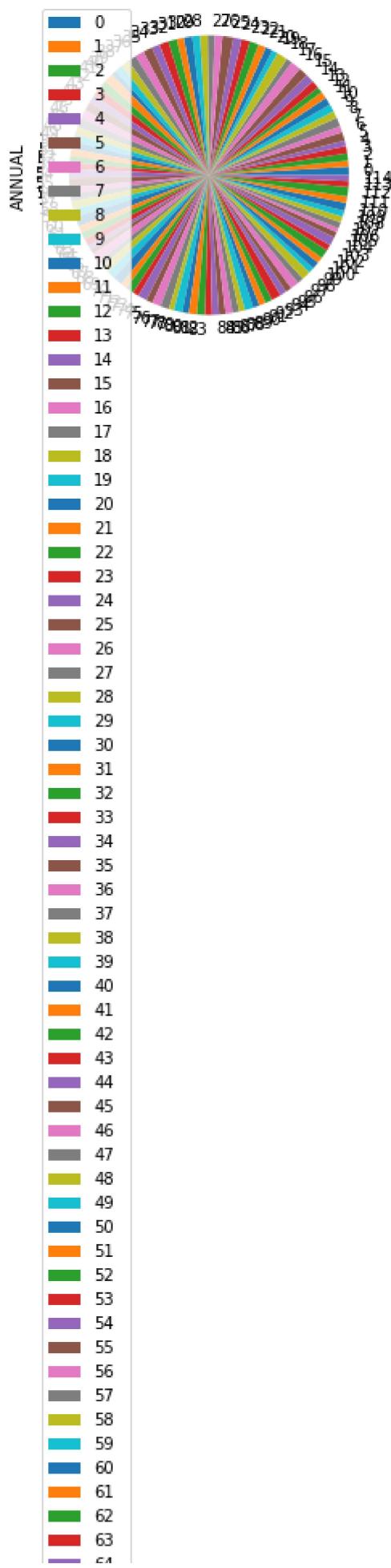
```
Out[16]: <AxesSubplot:
```

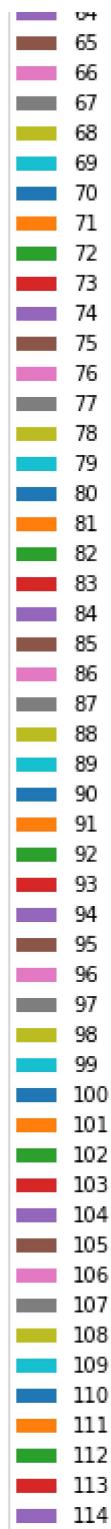


## Pie chart

```
In [17]: df.plot.pie(y='ANNUAL')
```

```
Out[17]: <AxesSubplot:ylabel='ANNUAL'>
```





## Scatter chart

```
In [18]: df.plot.scatter(x='SUBDIVISION' ,y='ANNUAL')
```

```
Out[18]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>
```



In [19]:

`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 [20]:

`df.describe()`

Out[20]:

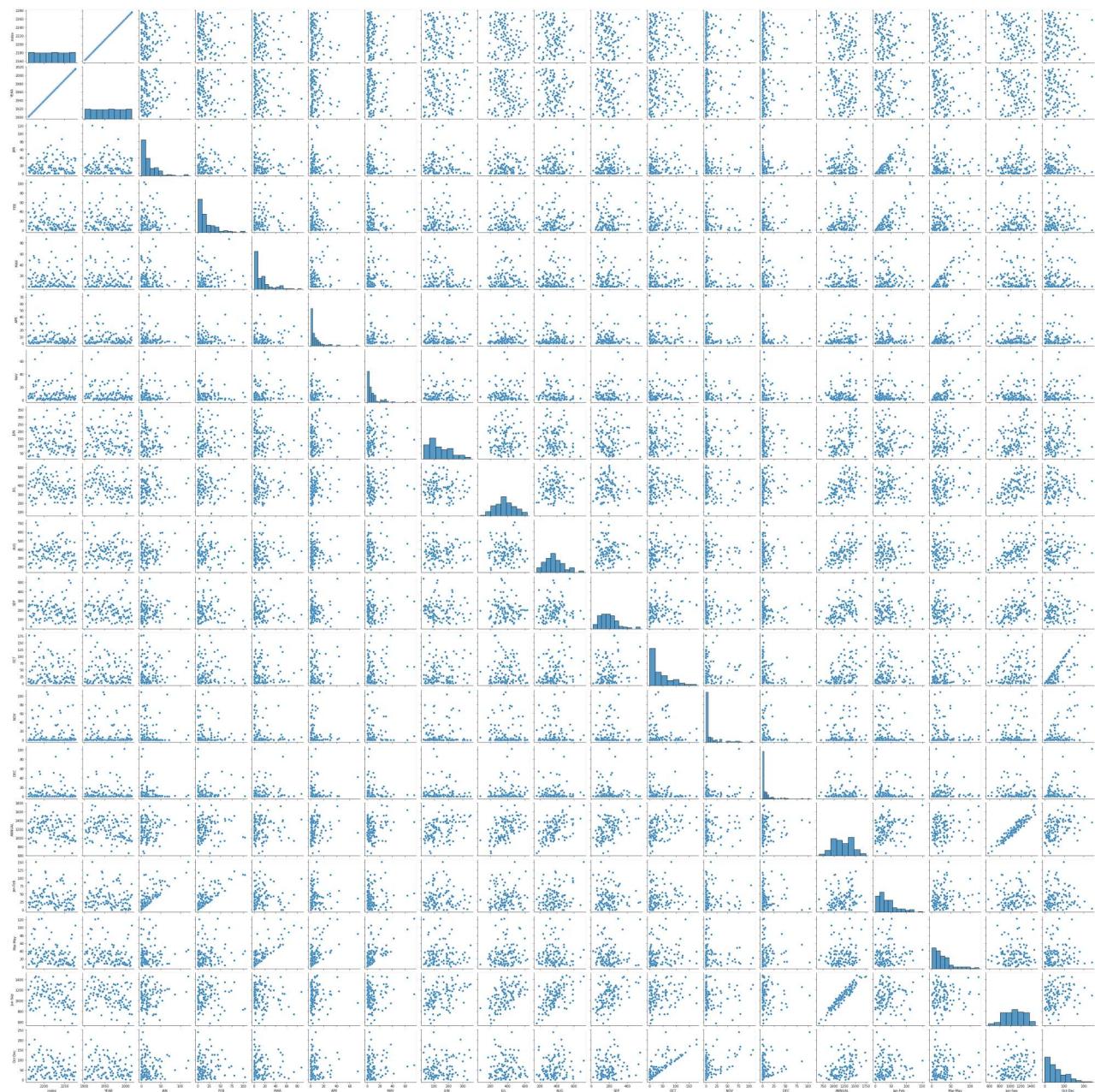
	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN
<b>count</b>	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000
<b>mean</b>	2219.000000	1958.000000	19.401739	18.693913	13.637391	7.188696	9.273043	141.029565
<b>std</b>	33.341666	33.341666	22.318347	20.795522	17.354996	10.473272	12.145379	79.359922
<b>min</b>	2162.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	26.300000
<b>25%</b>	2190.500000	1929.500000	2.200000	3.650000	1.150000	1.350000	2.100000	82.850000

	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN
<b>50%</b>	2219.000000	1958.000000	12.800000	11.300000	8.000000	3.200000	5.100000	118.500000
<b>75%</b>	2247.500000	1986.500000	29.650000	27.400000	18.650000	8.750000	10.500000	197.300000
<b>max</b>	2276.000000	2015.000000	120.700000	103.100000	87.300000	72.400000	74.200000	356.600000

## EDA AND VISUALIZATION

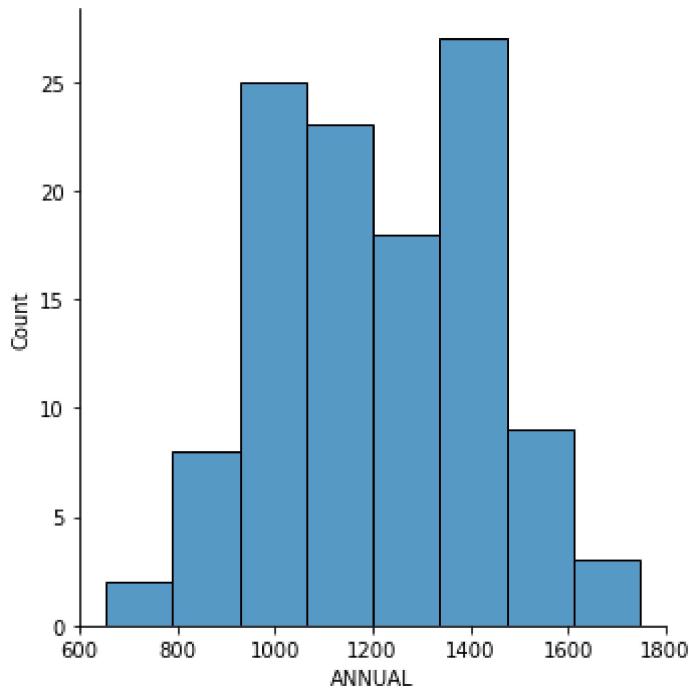
In [21]: `sns.pairplot(df)`

Out[21]: <seaborn.axisgrid.PairGrid at 0x2721c42c340>



In [22]: `sns.displot(df['ANNUAL'])`

Out[22]: &lt;seaborn.axisgrid.FacetGrid at 0x272278e0f70&gt;

In [23]:  
sns.heatmap(df.corr())

Out[23]: &lt;AxesSubplot:&gt;

