

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 uttar pradesh.csv")
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV
0	1012	EAST UTTAR PRADESH	1901	62.6	31.3	8.2	1.1	13.6	21.8	226.5	285.6	215.4	4.9	0.1
1	1013	EAST UTTAR PRADESH	1902	6.1	2.3	2.4	2.0	21.4	32.5	411.5	155.4	257.2	13.2	1.2
2	1014	EAST UTTAR PRADESH	1903	8.2	0.4	1.3	0.7	15.3	71.6	115.3	420.2	258.7	324.7	0.0
3	1015	EAST UTTAR PRADESH	1904	7.3	1.5	8.3	0.4	28.7	148.0	359.4	328.8	95.0	50.6	17.0
4	1016	EAST UTTAR PRADESH	1905	16.8	23.6	20.0	5.4	15.4	17.3	302.4	316.2	169.5	3.3	0.0
...
110	1122	EAST UTTAR PRADESH	2011	1.0	2.7	1.6	2.9	32.2	163.8	197.9	232.1	146.4	0.6	0.0
111	1123	EAST UTTAR PRADESH	2012	20.3	1.2	3.4	2.8	0.2	18.5	234.2	156.0	164.4	0.7	0.3
112	1124	EAST UTTAR PRADESH	2013	6.1	59.6	2.7	8.7	1.1	309.7	230.0	246.1	78.2	97.4	0.5
113	1125	EAST UTTAR PRADESH	2014	47.4	25.8	15.4	1.7	10.7	47.8	224.5	138.1	106.7	74.7	0.0
114	1126	EAST UTTAR PRADESH	2015	30.0	4.1	48.2	23.2	8.6	95.3	179.0	175.8	21.9	11.8	0.5

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	1012	EAST UTTAR PRADESH	1901	62.6	31.3	8.2	1.1	13.6	21.8	226.5	285.6	215.4	4.9	0.1
1	1013	EAST UTTAR PRADESH	1902	6.1	2.3	2.4	2.0	21.4	32.5	411.5	155.4	257.2	13.2	1.2
2	1014	EAST UTTAR PRADESH	1903	8.2	0.4	1.3	0.7	15.3	71.6	115.3	420.2	258.7	324.7	0.0
3	1015	EAST UTTAR PRADESH	1904	7.3	1.5	8.3	0.4	28.7	148.0	359.4	328.8	95.0	50.6	17.0
4	1016	EAST UTTAR PRADESH	1905	16.8	23.6	20.0	5.4	15.4	17.3	302.4	316.2	169.5	3.3	0.0
...
110	1122	EAST UTTAR PRADESH	2011	1.0	2.7	1.6	2.9	32.2	163.8	197.9	232.1	146.4	0.6	0.0
111	1123	EAST UTTAR PRADESH	2012	20.3	1.2	3.4	2.8	0.2	18.5	234.2	156.0	164.4	0.7	0.3
112	1124	EAST UTTAR PRADESH	2013	6.1	59.6	2.7	8.7	1.1	309.7	230.0	246.1	78.2	97.4	0.5
113	1125	EAST UTTAR PRADESH	2014	47.4	25.8	15.4	1.7	10.7	47.8	224.5	138.1	106.7	74.7	0.0
114	1126	EAST UTTAR PRADESH	2015	30.0	4.1	48.2	23.2	8.6	95.3	179.0	175.8	21.9	11.8	0.5

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	1012	EAST UTTAR PRADESH	1901	62.6	31.3	8.2	1.1	13.6	21.8	226.5	285.6	215.4	4.9	0.1
1	1013	EAST UTTAR PRADESH	1902	6.1	2.3	2.4	2.0	21.4	32.5	411.5	155.4	257.2	13.2	1.2
2	1014	EAST UTTAR PRADESH	1903	8.2	0.4	1.3	0.7	15.3	71.6	115.3	420.2	258.7	324.7	0.0

		index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV
3	1015		EAST UTTAR PRADESH	1904	7.3	1.5	8.3	0.4	28.7	148.0	359.4	328.8	95.0	50.6	17.0
4	1016		EAST UTTAR PRADESH	1905	16.8	23.6	20.0	5.4	15.4	17.3	302.4	316.2	169.5	3.3	0.0
...
110	1122		EAST UTTAR PRADESH	2011	1.0	2.7	1.6	2.9	32.2	163.8	197.9	232.1	146.4	0.6	0.0
111	1123		EAST UTTAR PRADESH	2012	20.3	1.2	3.4	2.8	0.2	18.5	234.2	156.0	164.4	0.7	0.3
112	1124		EAST UTTAR PRADESH	2013	6.1	59.6	2.7	8.7	1.1	309.7	230.0	246.1	78.2	97.4	0.5
113	1125		EAST UTTAR PRADESH	2014	47.4	25.8	15.4	1.7	10.7	47.8	224.5	138.1	106.7	74.7	0.0
114	1126		EAST UTTAR PRADESH	2015	30.0	4.1	48.2	23.2	8.6	95.3	179.0	175.8	21.9	11.8	0.5

115 rows × 20 columns

Data Cleaning and Data Preprocessing

describe()

In [5]:

```
df.describe()
```

Out[5]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN
count	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000
mean	1069.000000	1958.000000	16.012174	15.873913	8.907826	6.430435	17.211304	110.712174
std	33.341666	33.341666	14.659481	17.263607	11.906840	9.082731	14.095232	64.460799
min	1012.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.200000	17.300000
25%	1040.500000	1929.500000	3.450000	3.200000	0.700000	1.150000	6.550000	61.000000
50%	1069.000000	1958.000000	12.500000	8.800000	3.400000	3.800000	14.000000	99.600000
75%	1097.500000	1986.500000	24.400000	24.150000	12.050000	8.550000	25.900000	146.950000
max	1126.000000	2015.000000	62.600000	84.300000	63.700000	66.600000	68.700000	309.700000

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()
```

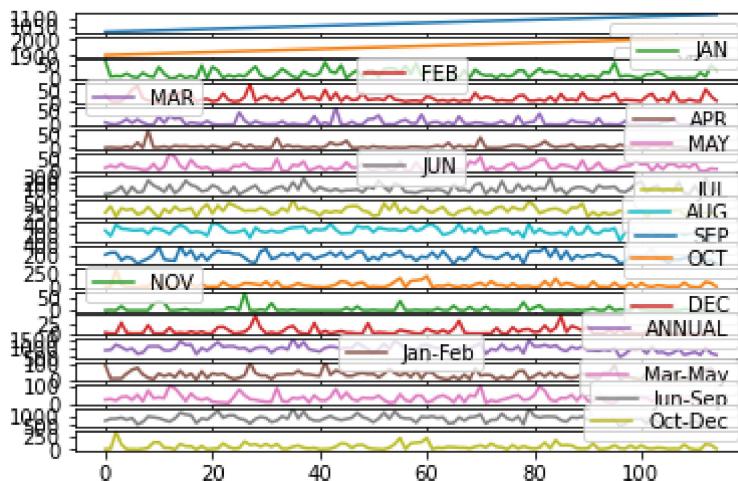
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 115 entries, 0 to 114
Data columns (total 20 columns):
 #   Column    Non-Null Count  Dtype  
--- 
 0   index     115 non-null   int64  
 1   SUBDIVISION 115 non-null  object  
 2   YEAR      115 non-null   int64  
 3   JAN        115 non-null   float64 
 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)
```

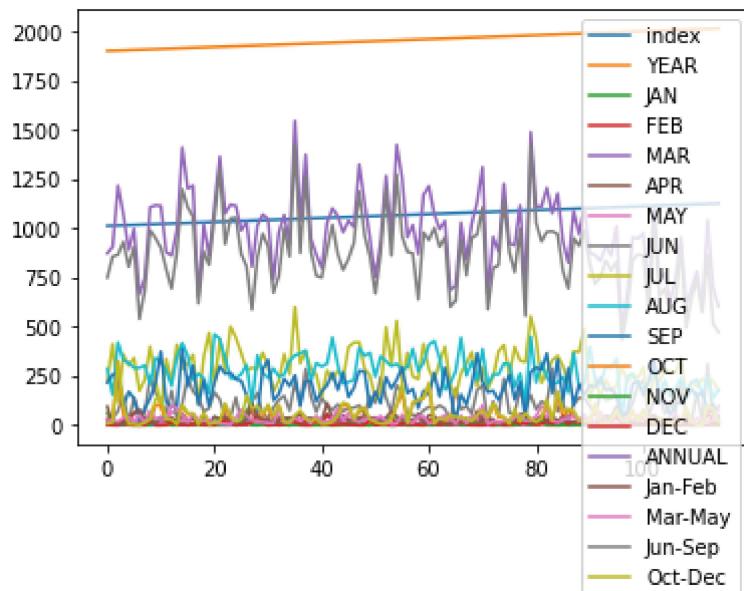
```
Out[11]: array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>], dtype=object)
```



Line chart

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

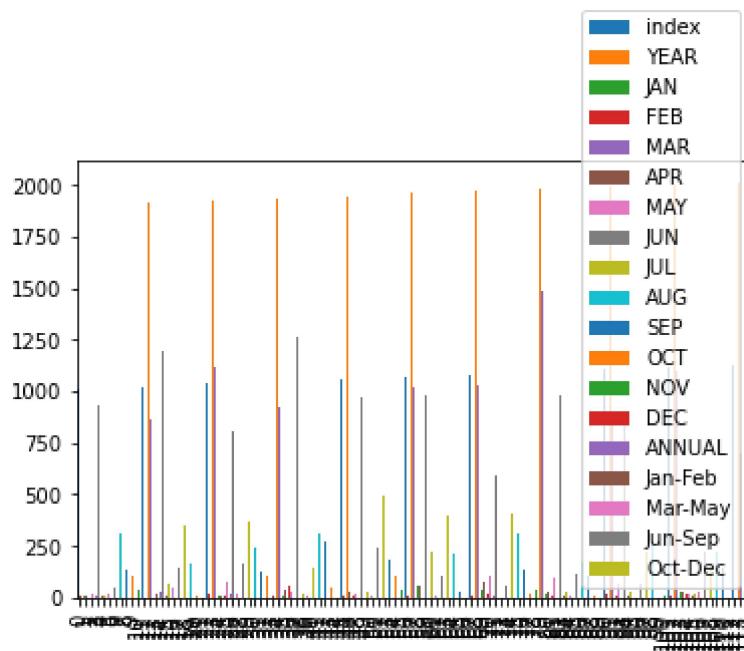
```
Out[12]: <AxesSubplot:>
```



Bar chart

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

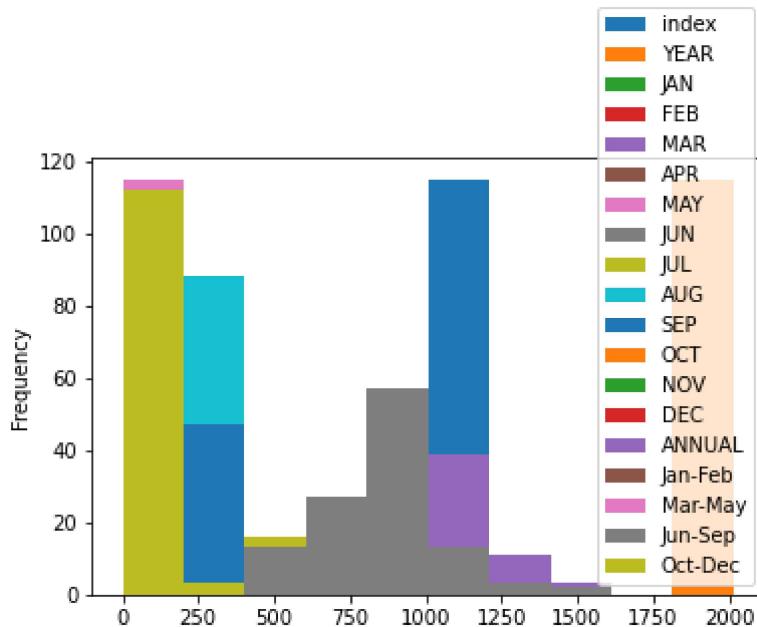
Out[13]: <AxesSubplot:>



Histogram

In [14]: `df.plot.hist()`

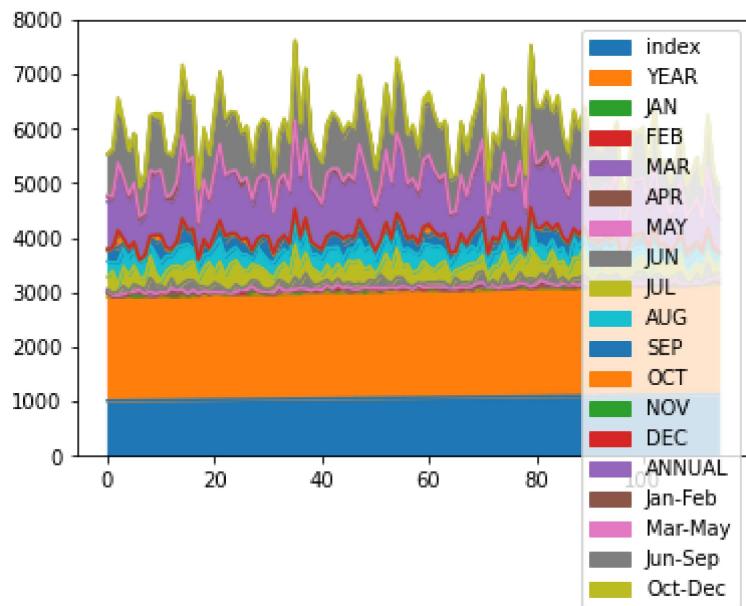
Out[14]: <AxesSubplot:ylabel='Frequency'>



Area chart

In [15]: `df.plot.area()`

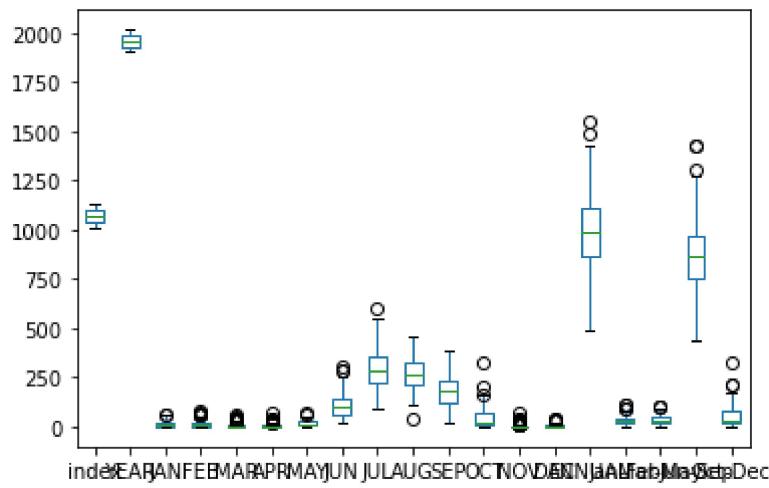
Out[15]: <AxesSubplot:>



Box chart

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

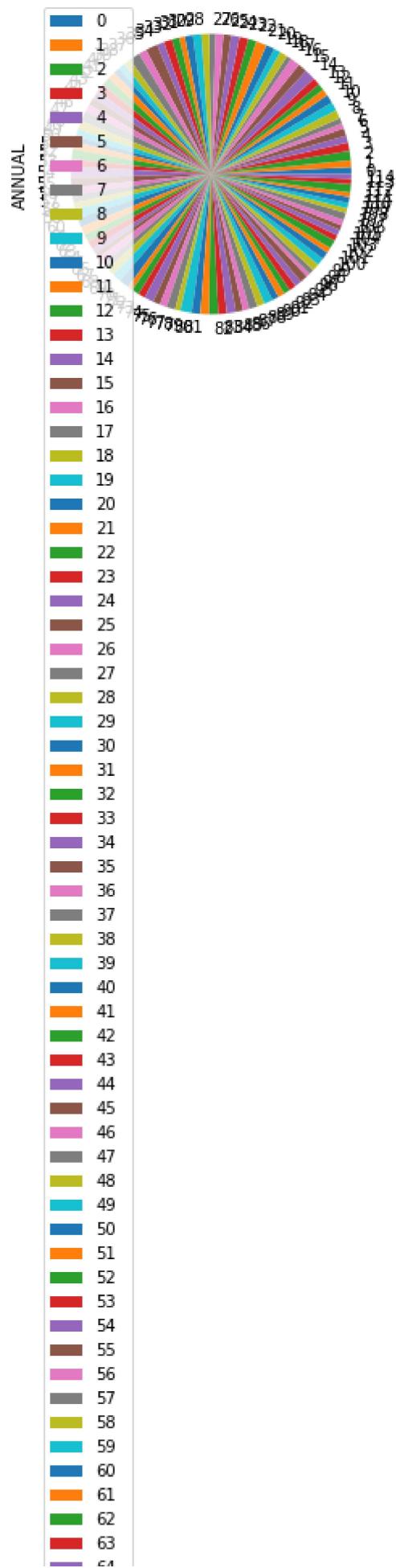
Out[16]: <AxesSubplot:>

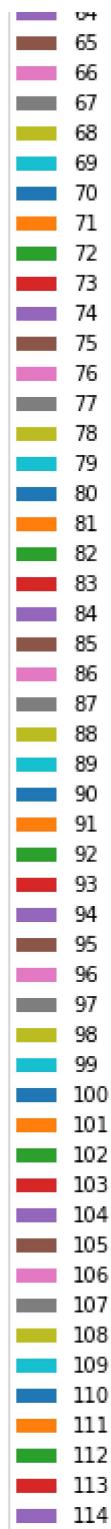


Pie chart

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

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

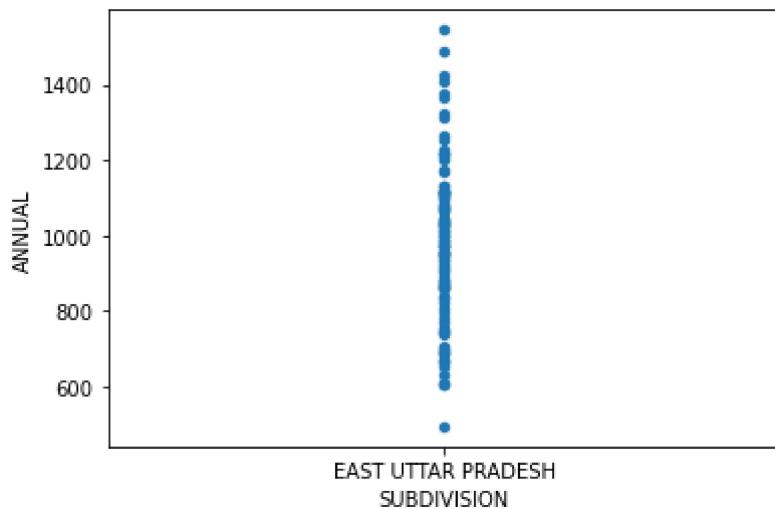




Scatter chart

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]:

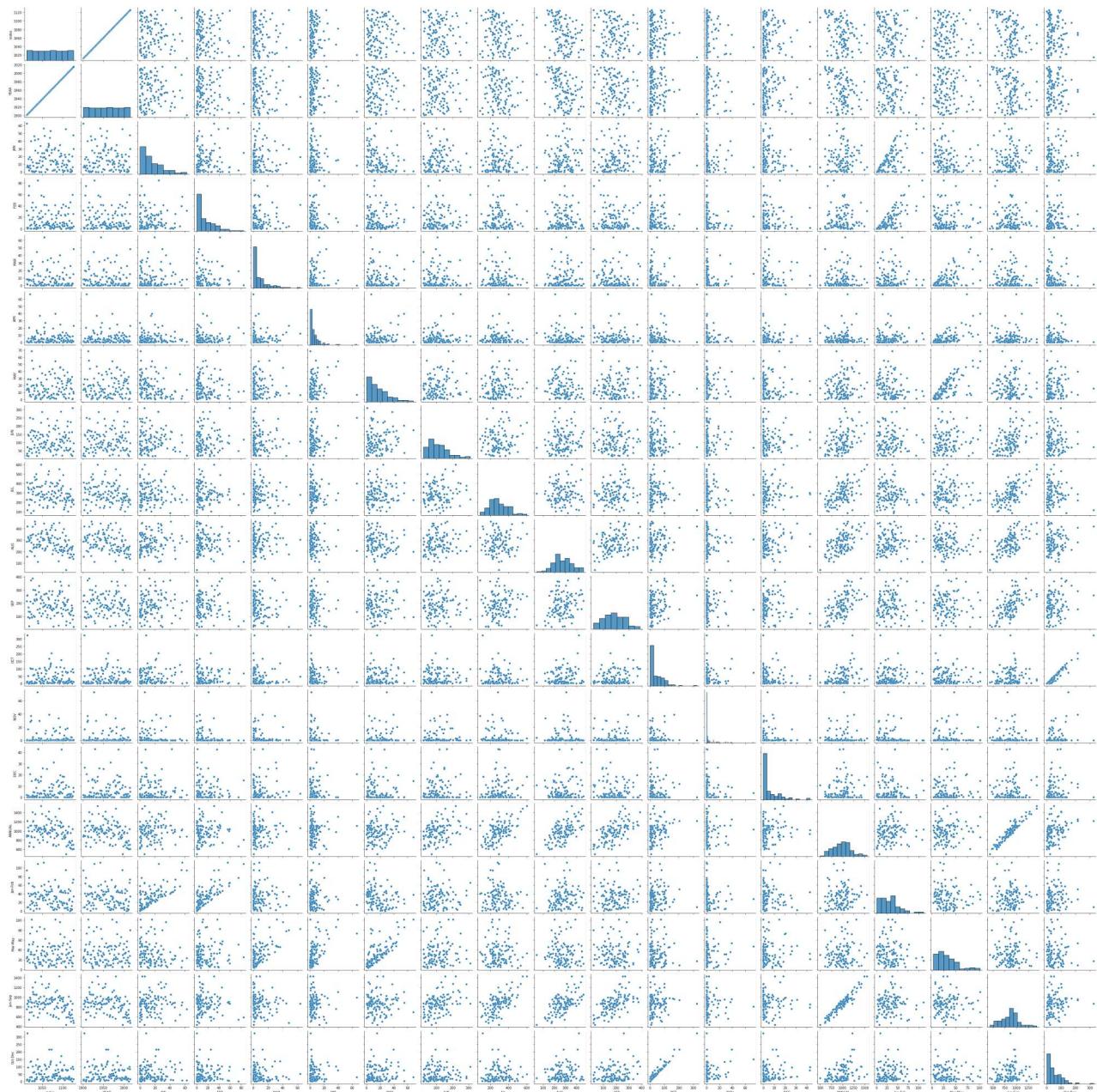
	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN
count	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000
mean	1069.000000	1958.000000	16.012174	15.873913	8.907826	6.430435	17.211304	110.712174
std	33.341666	33.341666	14.659481	17.263607	11.906840	9.082731	14.095232	64.460799
min	1012.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.200000	17.300000
25%	1040.500000	1929.500000	3.450000	3.200000	0.700000	1.150000	6.550000	61.000000

	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN
50%	1069.000000	1958.000000	12.500000	8.800000	3.400000	3.800000	14.000000	99.600000
75%	1097.500000	1986.500000	24.400000	24.150000	12.050000	8.550000	25.900000	146.950000
max	1126.000000	2015.000000	62.600000	84.300000	63.700000	66.600000	68.700000	309.700000

EDA AND VISUALIZATION

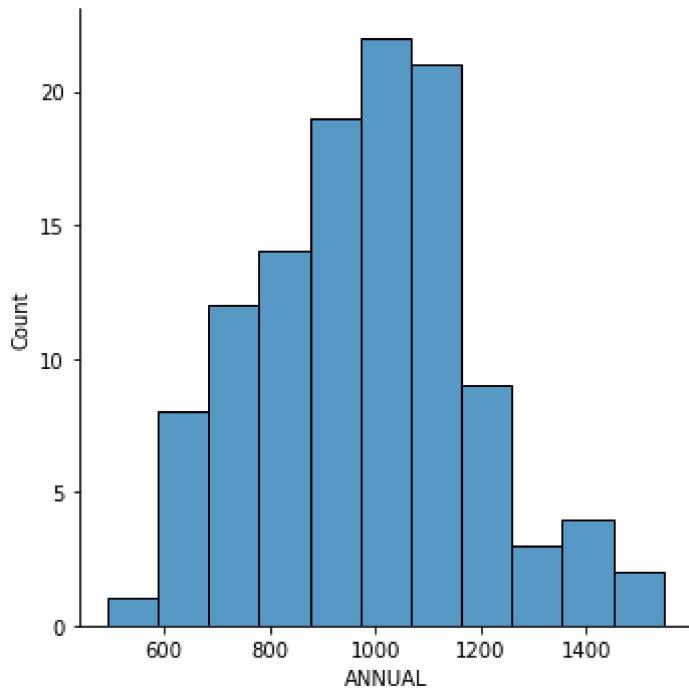
In [21]: `sns.pairplot(df)`

Out[21]: <seaborn.axisgrid.PairGrid at 0x22c7e743b80>



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

Out[22]: <seaborn.axisgrid.FacetGrid at 0x22c0ba2a130>

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

Out[23]: <AxesSubplot:>

