

# 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_haryana delhi _ chandigarh.csv")
df
```

Out[2]:

		index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	D
0	1357		HARYANA DELHI & CHANDIGARH	1901	35.4	28.9	11.1	0.0	5.1	13.2	126.4	151.5	10.5	2.0	0.0	
1	1358		HARYANA DELHI & CHANDIGARH	1902	0.0	0.7	2.9	10.2	15.8	74.6	149.3	97.1	59.8	9.3	0.0	
2	1359		HARYANA DELHI & CHANDIGARH	1903	14.7	0.5	2.3	0.5	8.5	8.6	151.6	138.2	97.7	4.0	0.0	
3	1360		HARYANA DELHI & CHANDIGARH	1904	7.6	0.7	48.0	0.5	29.3	34.3	109.7	162.9	102.3	1.5	10.4	2
4	1361		HARYANA DELHI & CHANDIGARH	1905	44.8	20.8	14.0	1.3	7.4	20.1	93.6	23.1	92.6	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
110	1467		HARYANA DELHI & CHANDIGARH	2011	0.7	26.7	6.9	8.9	28.7	94.4	85.0	127.3	133.1	0.0	0.0	
111	1468		HARYANA DELHI & CHANDIGARH	2012	8.2	0.2	0.1	11.8	3.8	5.3	68.1	196.6	90.7	2.4	0.6	
112	1469		HARYANA DELHI & CHANDIGARH	2013	21.1	52.2	5.3	3.3	1.4	62.1	96.5	161.9	42.8	10.9	1.7	
113	1470		HARYANA DELHI & CHANDIGARH	2014	13.0	17.3	26.8	7.5	20.3	25.9	72.3	34.8	67.3	10.5	0.2	
114	1471		HARYANA DELHI & CHANDIGARH	2015	12.4	6.6	71.8	34.8	8.4	43.7	130.3	89.2	32.1	3.7	2.3	

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	D
0	1357		HARYANA DELHI & CHANDIGARH	1901	35.4	28.9	11.1	0.0	5.1	13.2	126.4	151.5	10.5	2.0	0.0	
1	1358		HARYANA DELHI & CHANDIGARH	1902	0.0	0.7	2.9	10.2	15.8	74.6	149.3	97.1	59.8	9.3	0.0	
2	1359		HARYANA DELHI & CHANDIGARH	1903	14.7	0.5	2.3	0.5	8.5	8.6	151.6	138.2	97.7	4.0	0.0	
3	1360		HARYANA DELHI & CHANDIGARH	1904	7.6	0.7	48.0	0.5	29.3	34.3	109.7	162.9	102.3	1.5	10.4	2
4	1361		HARYANA DELHI & CHANDIGARH	1905	44.8	20.8	14.0	1.3	7.4	20.1	93.6	23.1	92.6	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
110	1467		HARYANA DELHI & CHANDIGARH	2011	0.7	26.7	6.9	8.9	28.7	94.4	85.0	127.3	133.1	0.0	0.0	
111	1468		HARYANA DELHI & CHANDIGARH	2012	8.2	0.2	0.1	11.8	3.8	5.3	68.1	196.6	90.7	2.4	0.6	
112	1469		HARYANA DELHI & CHANDIGARH	2013	21.1	52.2	5.3	3.3	1.4	62.1	96.5	161.9	42.8	10.9	1.7	
113	1470		HARYANA DELHI & CHANDIGARH	2014	13.0	17.3	26.8	7.5	20.3	25.9	72.3	34.8	67.3	10.5	0.2	
114	1471		HARYANA DELHI & CHANDIGARH	2015	12.4	6.6	71.8	34.8	8.4	43.7	130.3	89.2	32.1	3.7	2.3	

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	DEC	
		0	HARYANA DELHI & CHANDIGARH	1357	1901	35.4	28.9	11.1	0.0	5.1	13.2	126.4	151.5	10.5	2.0	0.0	
		1	HARYANA DELHI & CHANDIGARH	1358	1902	0.0	0.7	2.9	10.2	15.8	74.6	149.3	97.1	59.8	9.3	0.0	
		2	HARYANA DELHI & CHANDIGARH	1359	1903	14.7	0.5	2.3	0.5	8.5	8.6	151.6	138.2	97.7	4.0	0.0	
		3	HARYANA DELHI & CHANDIGARH	1360	1904	7.6	0.7	48.0	0.5	29.3	34.3	109.7	162.9	102.3	1.5	10.4	2
		4	HARYANA DELHI & CHANDIGARH	1361	1905	44.8	20.8	14.0	1.3	7.4	20.1	93.6	23.1	92.6	0.0	0.0	
		...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
		110	HARYANA DELHI & CHANDIGARH	1467	2011	0.7	26.7	6.9	8.9	28.7	94.4	85.0	127.3	133.1	0.0	0.0	
		111	HARYANA DELHI & CHANDIGARH	1468	2012	8.2	0.2	0.1	11.8	3.8	5.3	68.1	196.6	90.7	2.4	0.6	
		112	HARYANA DELHI & CHANDIGARH	1469	2013	21.1	52.2	5.3	3.3	1.4	62.1	96.5	161.9	42.8	10.9	1.7	
		113	HARYANA DELHI & CHANDIGARH	1470	2014	13.0	17.3	26.8	7.5	20.3	25.9	72.3	34.8	67.3	10.5	0.2	
		114	HARYANA DELHI & CHANDIGARH	1471	2015	12.4	6.6	71.8	34.8	8.4	43.7	130.3	89.2	32.1	3.7	2.3	

115 rows × 20 columns



## Data Cleaning and Data Preprocessing

### describe()

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

	<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>	1414.000000	1958.000000	16.889565	17.433913	12.935652	7.633913	14.533913	48.626087
<b>std</b>	33.341666	33.341666	15.514478	18.893422	15.251840	12.847533	15.900347	34.183448
<b>min</b>	1357.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	2.100000
<b>25%</b>	1385.500000	1929.500000	3.550000	2.250000	2.100000	0.800000	3.700000	24.600000
<b>50%</b>	1414.000000	1958.000000	14.300000	12.100000	7.200000	2.800000	7.900000	43.700000
<b>75%</b>	1442.500000	1986.500000	25.150000	27.850000	17.700000	8.750000	20.700000	66.600000
<b>max</b>	1471.000000	2015.000000	66.500000	91.000000	71.800000	82.500000	72.900000	193.500000

**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()`

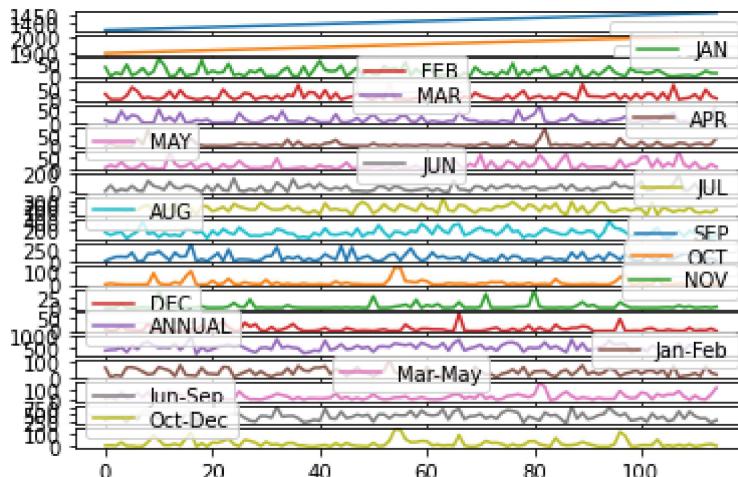
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 115 entries, 0 to 114
Data columns (total 20 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   index       115 non-null    int64  
 1   SUBDIVISION 115 non-null    object  
 2   YEAR        115 non-null    int64  
 3   JAN         115 non-null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

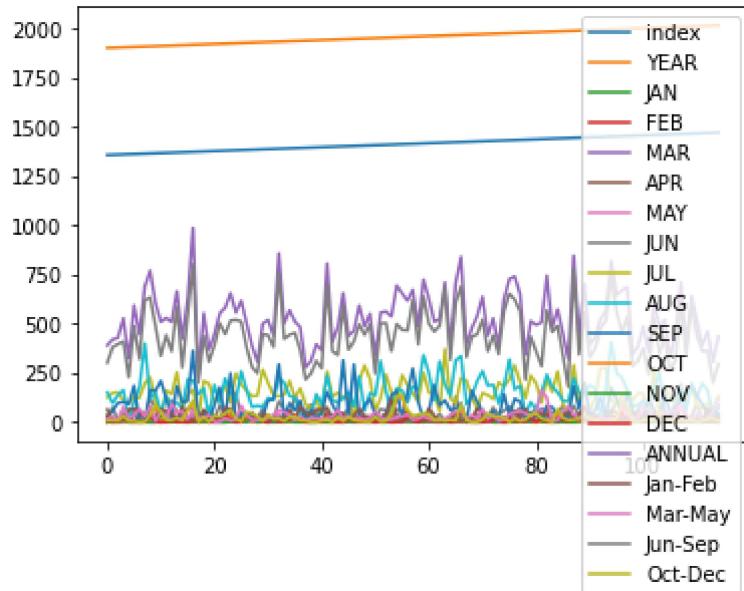
```
Out[11]: array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
   <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
   <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
   <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
   <AxesSubplot:>,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



# Line chart

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

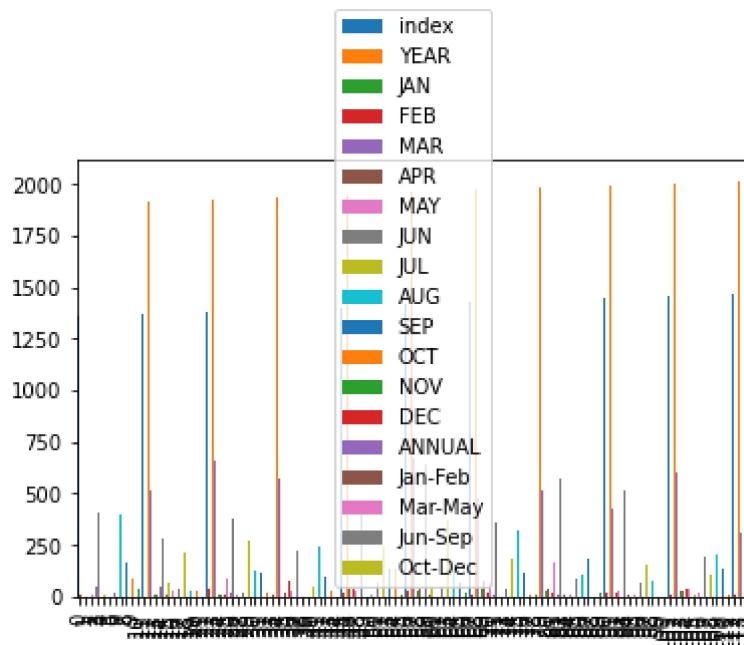
```
Out[12]: <AxesSubplot:>
```



# Bar chart

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

```
Out[13]: <AxesSubplot:>
```

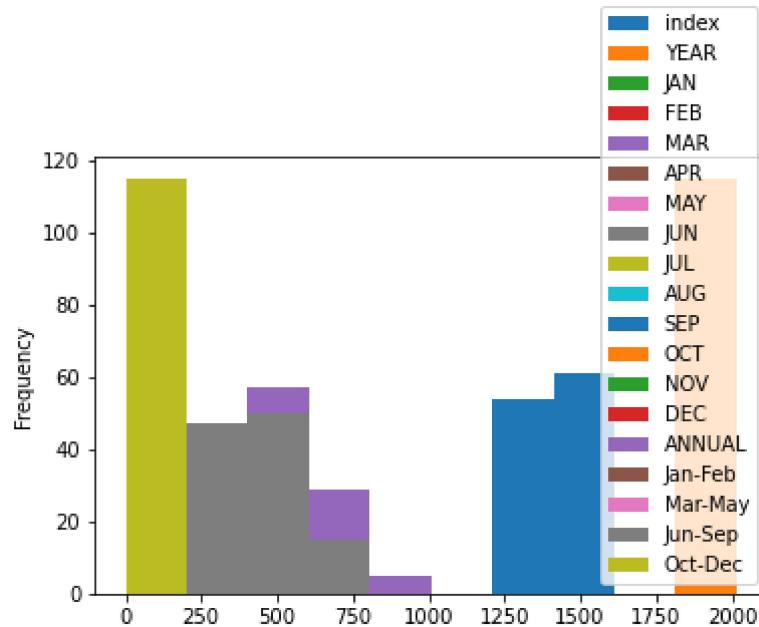


# Histogram

In [14]:

df.plot.hist()

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

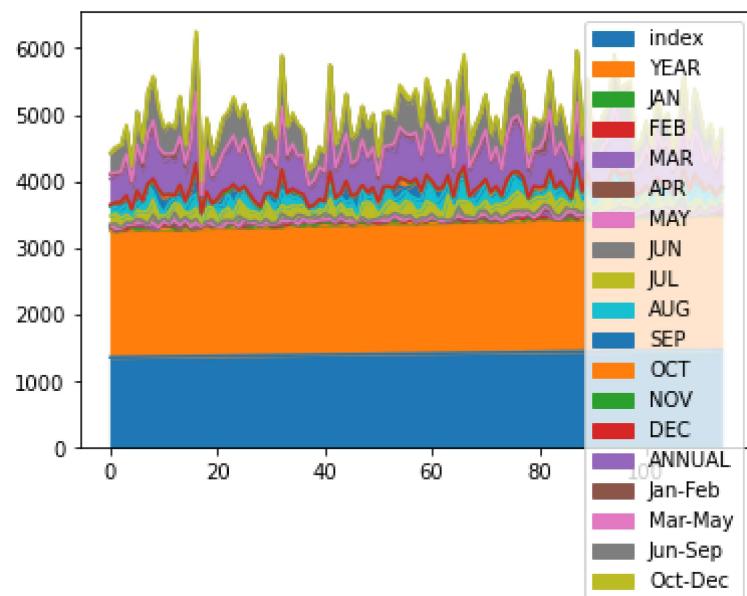


# Area chart

In [15]:

df.plot.area()

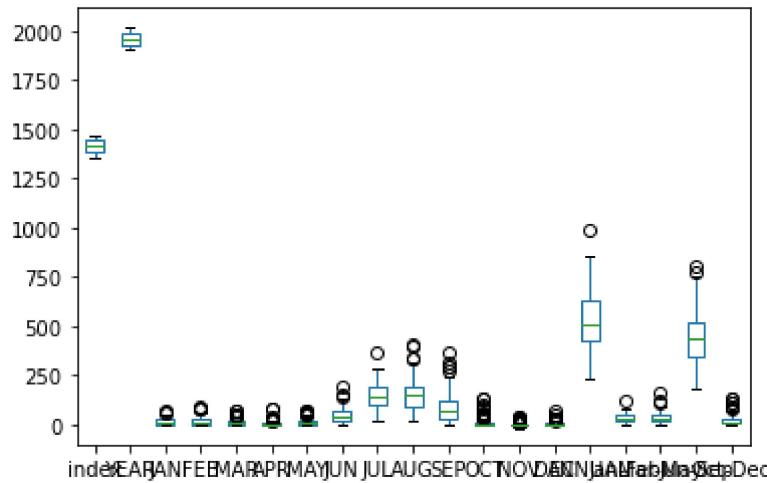
Out[15]: &lt;AxesSubplot:&gt;



# Box chart

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

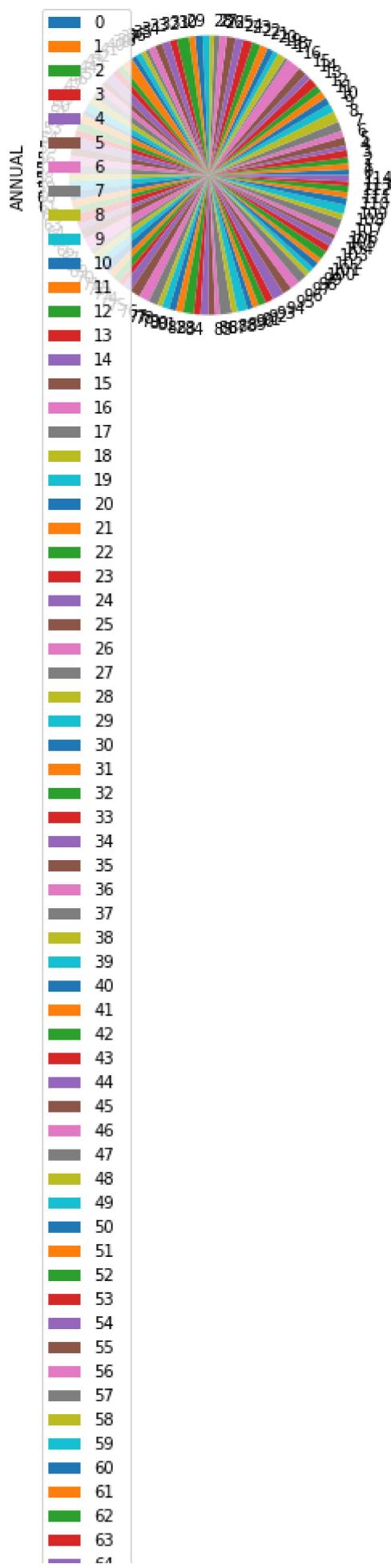
```
Out[16]: <AxesSubplot:
```

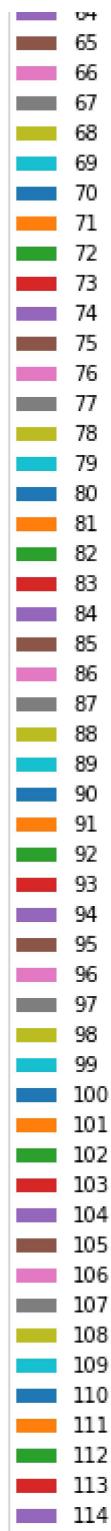


## Pie chart

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

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

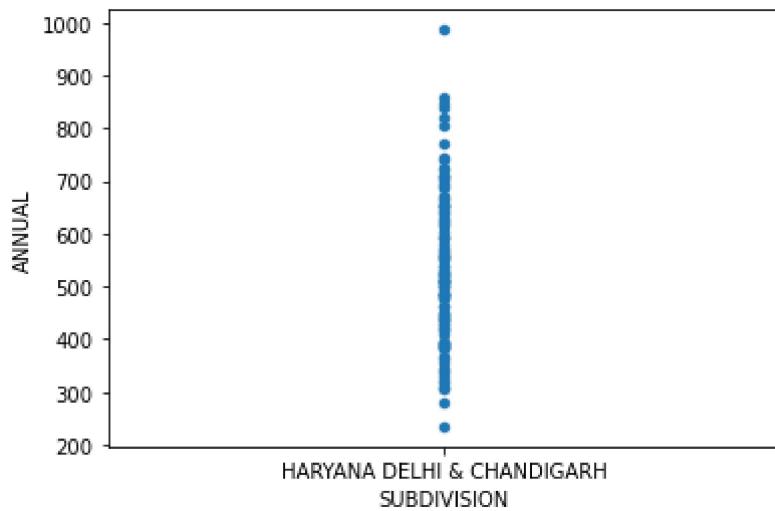




## Scatter chart

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]:

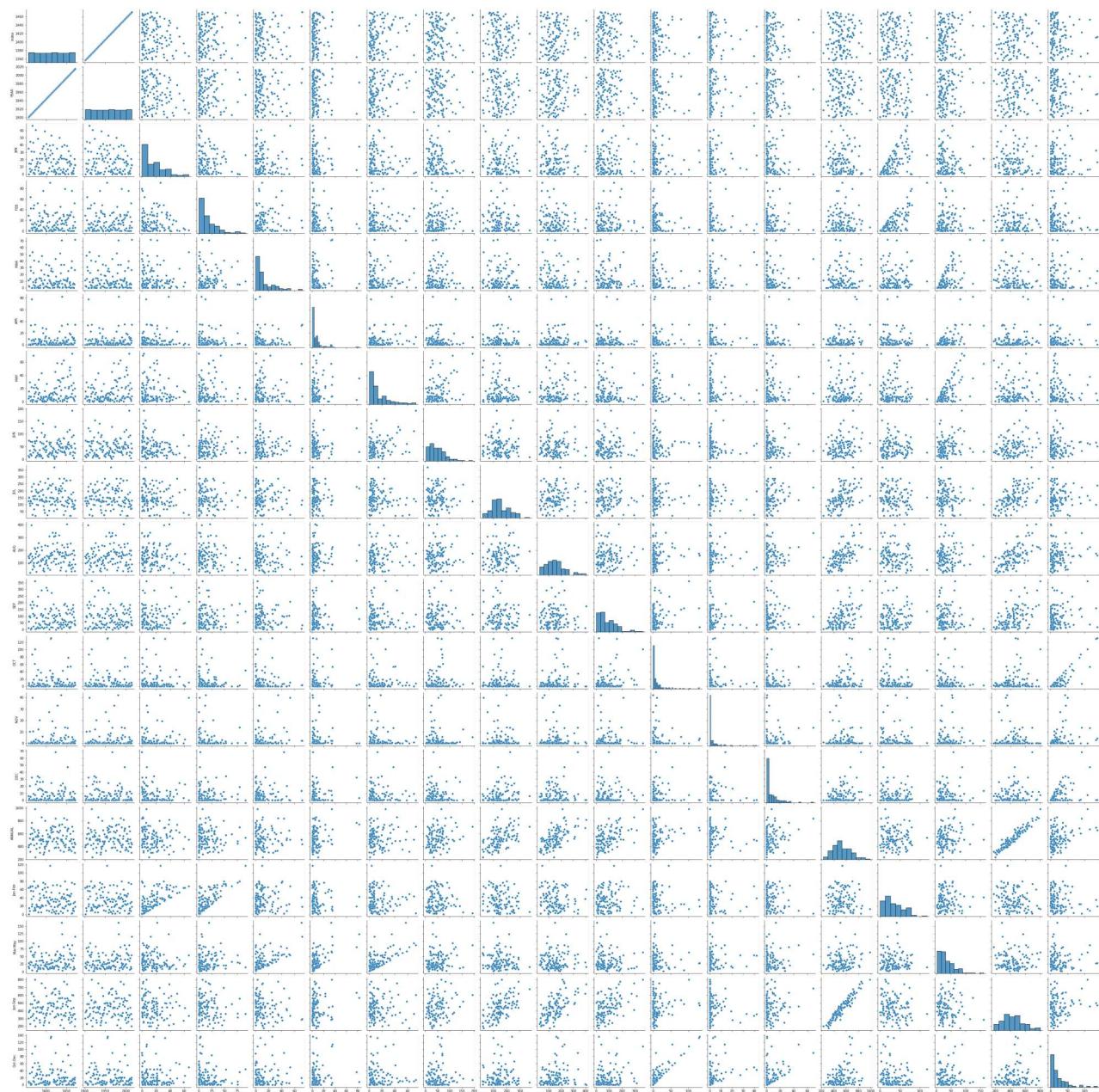
	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN
<b>count</b>	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000
<b>mean</b>	1414.000000	1958.000000	16.889565	17.433913	12.935652	7.633913	14.533913	48.626087
<b>std</b>	33.341666	33.341666	15.514478	18.893422	15.251840	12.847533	15.900347	34.183448
<b>min</b>	1357.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	2.100000
<b>25%</b>	1385.500000	1929.500000	3.550000	2.250000	2.100000	0.800000	3.700000	24.600000

	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN
<b>50%</b>	1414.000000	1958.000000	14.300000	12.100000	7.200000	2.800000	7.900000	43.700000
<b>75%</b>	1442.500000	1986.500000	25.150000	27.850000	17.700000	8.750000	20.700000	66.600000
<b>max</b>	1471.000000	2015.000000	66.500000	91.000000	71.800000	82.500000	72.900000	193.500000

## EDA AND VISUALIZATION

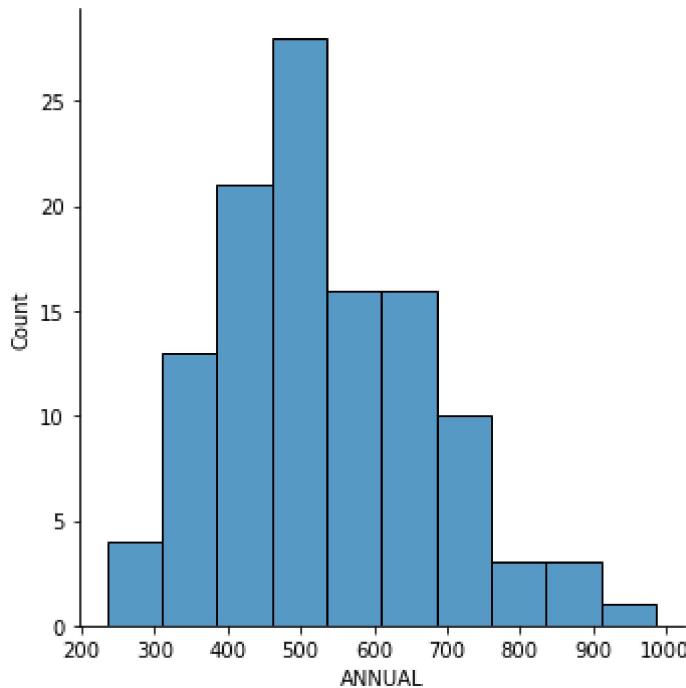
In [21]: `sns.pairplot(df)`

Out[21]: <seaborn.axisgrid.PairGrid at 0x2bf7d055e80>



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

Out[22]: &lt;seaborn.axisgrid.FacetGrid at 0x2bf0a500340&gt;

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

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

