Import Libraries

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:

```
df=pd.read_csv(r"c:\Users\user\Downloads\FP2_RainFall\rainfall.csv")[3544:3657]
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SE
3544	3544	COASTAL KARNATAKA	1903	0.7	0.0	0.0	4.1	202.8	536.5	1405.5	593.8	304
3545	3545	COASTAL KARNATAKA	1904	2.4	0.0	4.8	23.7	93.2	1108.2	1070.0	465.6	245
3546	3546	COASTAL KARNATAKA	1905	0.0	0.2	0.0	6.4	83.1	767.3	777.3	586.9	172
3547	3547	COASTAL KARNATAKA	1906	23.0	0.0	0.0	0.5	29.8	593.6	1173.4	535.0	273
3548	3548	COASTAL KARNATAKA	1907	7.0	0.0	2.7	91.3	19.0	845.9	1166.3	1026.1	223
...	...	...	...	...	...	...	...	...	...	...	...	...
3652	3652	COASTAL KARNATAKA	2011	4.8	3.8	8.7	66.1	49.3	1018.4	1080.5	861.3	545
3653	3653	COASTAL KARNATAKA	2012	NaN	11.4	5.1	77.0	22.9	650.9	754.6	1027.6	382
3654	3654	COASTAL KARNATAKA	2013	2.4	19.6	19.0	28.5	100.4	1153.0	1515.3	680.2	379
3655	3655	COASTAL KARNATAKA	2014	0.0	0.3	1.9	40.5	181.9	507.0	1155.4	1121.0	379
3656	3656	COASTAL KARNATAKA	2015	1.4	1.0	32.3	72.2	150.3	735.3	930.9	575.2	260

113 rows × 20 columns

Data Cleaning and Preprocessing

In [3]:

```
df.dropna()
```

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SE
3544	3544	COASTAL KARNATAKA	1903	0.7	0.0	0.0	4.1	202.8	536.5	1405.5	593.8	304
3545	3545	COASTAL KARNATAKA	1904	2.4	0.0	4.8	23.7	93.2	1108.2	1070.0	465.6	245
3546	3546	COASTAL KARNATAKA	1905	0.0	0.2	0.0	6.4	83.1	767.3	777.3	586.9	172
3547	3547	COASTAL KARNATAKA	1906	23.0	0.0	0.0	0.5	29.8	593.6	1173.4	535.0	273
3548	3548	COASTAL KARNATAKA	1907	7.0	0.0	2.7	91.3	19.0	845.9	1166.3	1026.1	223
...	...	...	...	...	...	...	...	...	...	...	...	
3651	3651	COASTAL KARNATAKA	2010	14.4	0.4	3.5	62.2	80.2	682.7	1200.2	637.5	468
3652	3652	COASTAL KARNATAKA	2011	4.8	3.8	8.7	66.1	49.3	1018.4	1080.5	861.3	545
3654	3654	COASTAL KARNATAKA	2013	2.4	19.6	19.0	28.5	100.4	1153.0	1515.3	680.2	379
3655	3655	COASTAL KARNATAKA	2014	0.0	0.3	1.9	40.5	181.9	507.0	1155.4	1121.0	379
3656	3656	COASTAL KARNATAKA	2015	1.4	1.0	32.3	72.2	150.3	735.3	930.9	575.2	260

112 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-Fe
b',
       'Mar-May', 'Jun-Sep', 'Oct-Dec'],
      dtype='object')
```

In [5]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113 entries, 3544 to 3656
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   index        113 non-null    int64  
 1   SUBDIVISION  113 non-null    object 
 2   YEAR         113 non-null    int64  
 3   JAN          112 non-null    float64
 4   FEB          113 non-null    float64
 5   MAR          113 non-null    float64
 6   APR          113 non-null    float64
 7   MAY          113 non-null    float64
 8   JUN          113 non-null    float64
 9   JUL          113 non-null    float64
 10  AUG          113 non-null    float64
 11  SEP          113 non-null    float64
 12  OCT          113 non-null    float64
 13  NOV          113 non-null    float64
 14  DEC          113 non-null    float64
 15  ANNUAL       112 non-null    float64
 16  Jan-Feb      112 non-null    float64
 17  Mar-May      113 non-null    float64
 18  Jun-Sep      113 non-null    float64
 19  Oct-Dec      113 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 17.8+ KB
```

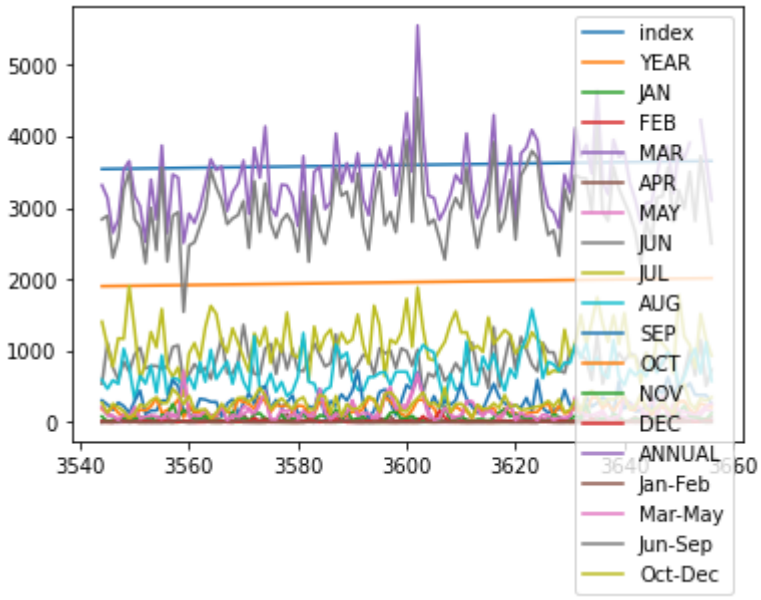
Line Chart

In [6]:

```
df.plot.line()
```

Out[6]:

<AxesSubplot:>



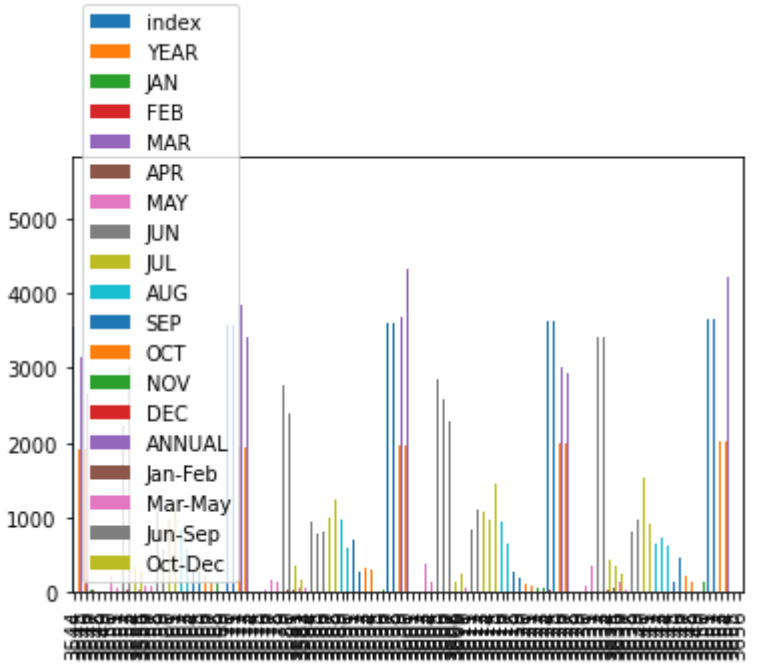
Bar chart

In [7]:

```
df.plot.bar()
```

Out[7]:

<AxesSubplot:>



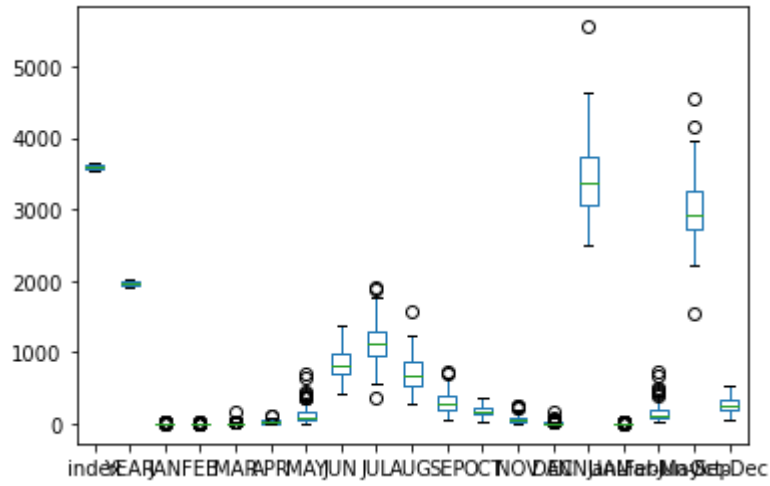
Box chart

In [8]:

```
df.plot.box()
```

Out[8]:

<AxesSubplot:>



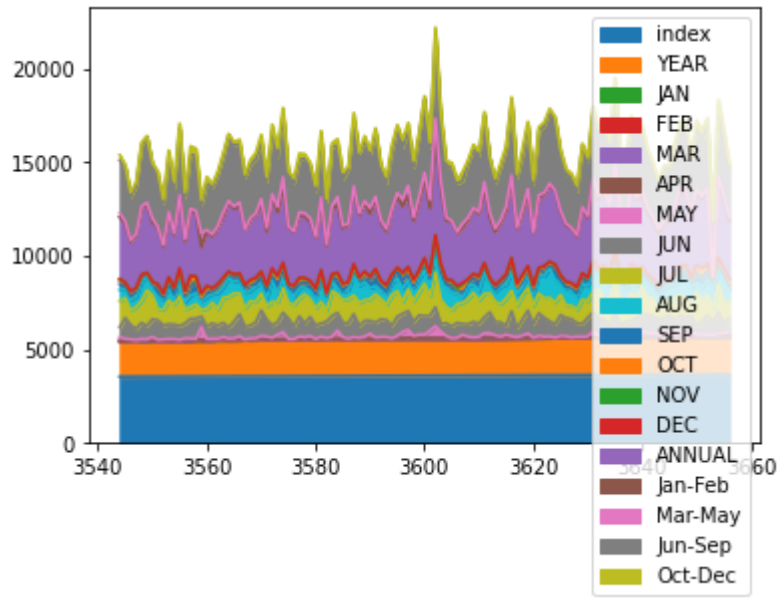
Area Chart

In [9]:

```
df.plot.area()
```

Out[9]:

<AxesSubplot:>



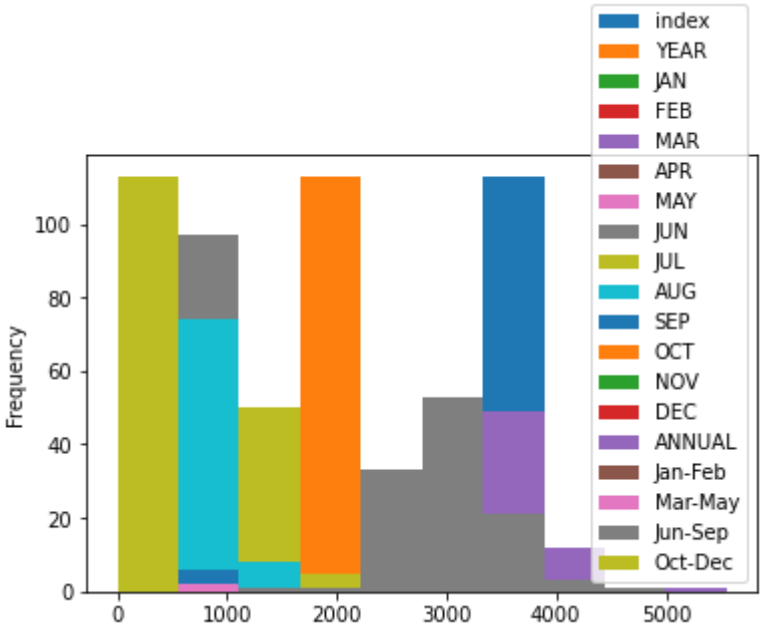
Histogram

In [10]:

```
df.plot.hist()
```

Out[10]:

```
<AxesSubplot:ylabel='Frequency'>
```



pie chart

In [11]:

```
df.plot.pie(y='ANNUAL')
```

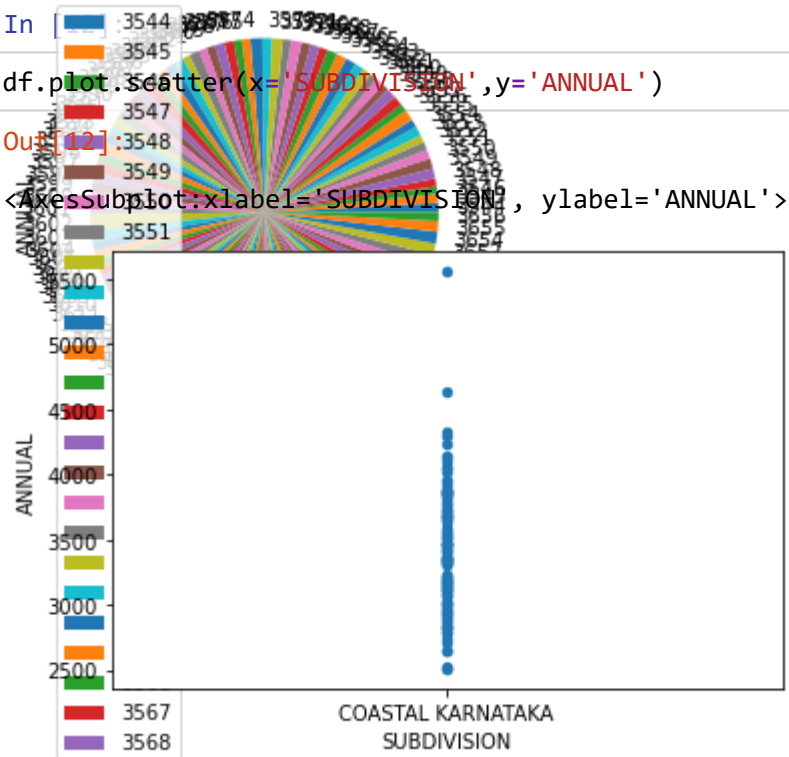
Out[11]:

```
<AxesSubplot:ylabel='ANNUAL'>
```

Scatter chart

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

Out[12]:
```
<AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>
```



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

Out[13]:

	index	YEAR	JAN	FEB	MAR	APR	MAY
count	113.00000	113.00000	112.000000	113.000000	113.000000	113.000000	113.000000
mean	3600.00000	1959.00000	1.927679	1.537168	6.331858	30.909735	123.755752
std	32.76431	32.76431	4.254477	4.769082	16.715034	24.184030	126.280047
min	3544.00000	1903.00000	0.000000	0.000000	0.000000	0.000000	8.400000
25%	3572.00000	1931.00000	0.000000	0.000000	0.200000	11.400000	40.200000
50%	3600.00000	1959.00000	0.100000	0.000000	1.400000	24.700000	80.200000
75%	3628.00000	1987.00000	1.925000	0.500000	5.700000	44.800000	164.000000
max	3656.00000	2015.00000	23.000000	29.800000	161.400000	110.100000	699.500000

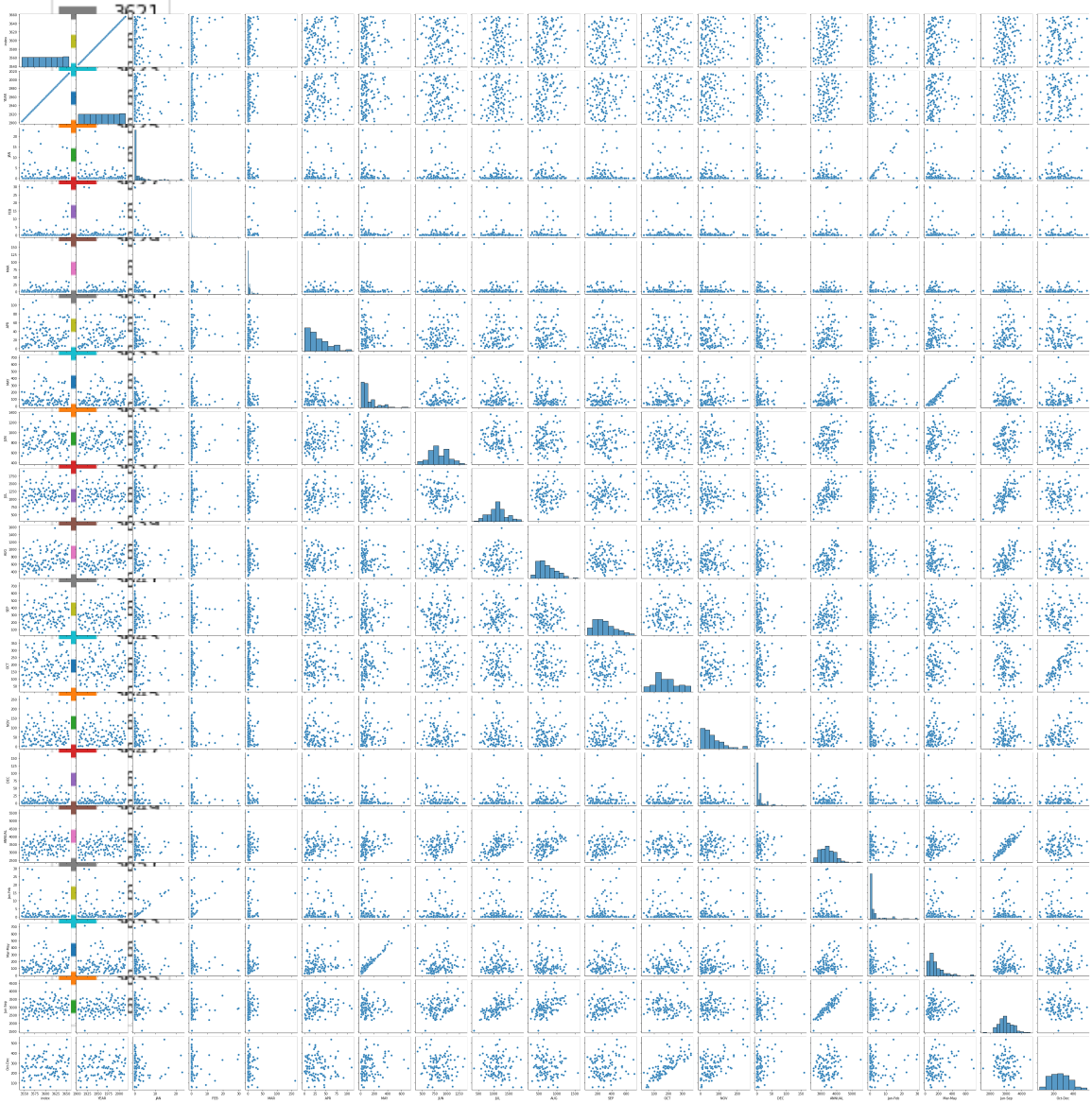
EDA and Visualization

In [14]:

```
sns.pairplot(df)
```

Out[14]:

```
<seaborn.axisgrid.PairGrid at 0x241ff91cc40>
```

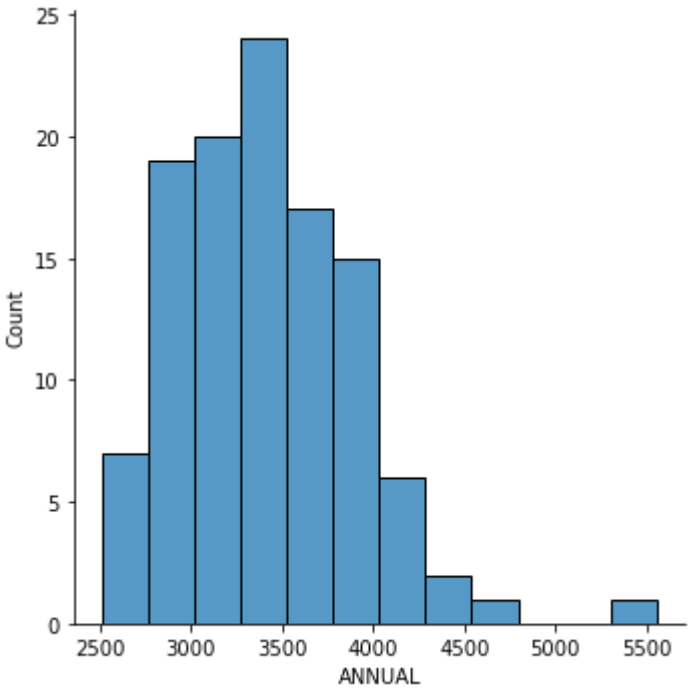


In [15]:

```
sns.displot(df['ANNUAL'])
```

Out[15]:

```
<seaborn.axisgrid.FacetGrid at 0x24184ca8ca0>
```

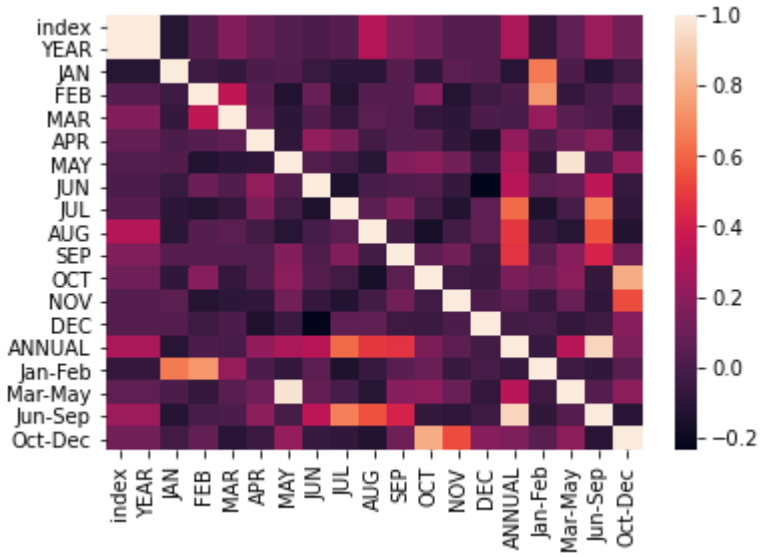


In [16]:

```
sns.heatmap(df.corr())
```

Out[16]:

```
<AxesSubplot:>
```



In [ ]: