

Importing Libraries

In [1]:

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

Importing Datasets

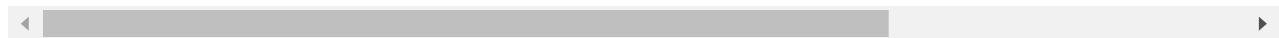
In [2]:

```
df=pd.read_csv("rainfall_assam _ meghalaya.csv")
df
```

Out[2]:

| | index | SUBDIVISION | YEAR | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV |
|-----|-------|-------------------|------|------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0 | 207 | ASSAM & MEGHALAYA | 1901 | 27.1 | 19.5 | 30.6 | 223.0 | 207.0 | 524.9 | 430.6 | 464.1 | 291.4 | 163.7 | 115.6 |
| 1 | 208 | ASSAM & MEGHALAYA | 1902 | 9.3 | 10.2 | 105.6 | 350.0 | 262.1 | 620.7 | 510.8 | 536.0 | 441.3 | 97.0 | 7.8 |
| 2 | 209 | ASSAM & MEGHALAYA | 1903 | 19.9 | 25.4 | 103.6 | 140.6 | 206.6 | 607.4 | 362.7 | 551.9 | 306.4 | 159.5 | 59.3 |
| 3 | 210 | ASSAM & MEGHALAYA | 1904 | 11.1 | 56.1 | 51.9 | 457.1 | 375.2 | 385.7 | 477.6 | 438.8 | 245.9 | 115.9 | 46.4 |
| 4 | 211 | ASSAM & MEGHALAYA | 1905 | 19.9 | 16.9 | 137.9 | 213.0 | 275.5 | 521.7 | 439.1 | 649.1 | 276.0 | 200.0 | 16.8 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 110 | 317 | ASSAM & MEGHALAYA | 2011 | 11.1 | 11.4 | 109.0 | 92.1 | 238.3 | 316.0 | 395.8 | 302.6 | 221.6 | 30.2 | 11.9 |
| 111 | 318 | ASSAM & MEGHALAYA | 2012 | 15.2 | 6.9 | 28.8 | 279.1 | 185.8 | 729.7 | 444.3 | 289.2 | 411.6 | 199.4 | 17.1 |
| 112 | 319 | ASSAM & MEGHALAYA | 2013 | 1.1 | 9.6 | 44.0 | 112.8 | 346.7 | 286.2 | 367.8 | 289.7 | 229.3 | 126.3 | 1.0 |
| 113 | 320 | ASSAM & MEGHALAYA | 2014 | 2.0 | 28.3 | 29.3 | 51.5 | 351.1 | 426.4 | 374.4 | 484.6 | 420.2 | 35.0 | 3.0 |
| 114 | 321 | ASSAM & MEGHALAYA | 2015 | 13.4 | 15.5 | 37.5 | 250.9 | 332.5 | 558.5 | 300.1 | 590.9 | 279.9 | 62.6 | 14.0 |

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 | 207 | ASSAM & MEGHALAYA | 1901 | 27.1 | 19.5 | 30.6 | 223.0 | 207.0 | 524.9 | 430.6 | 464.1 | 291.4 | 163.7 | 115.6 |
| 1 | 208 | ASSAM & MEGHALAYA | 1902 | 9.3 | 10.2 | 105.6 | 350.0 | 262.1 | 620.7 | 510.8 | 536.0 | 441.3 | 97.0 | 7.8 |
| 2 | 209 | ASSAM & MEGHALAYA | 1903 | 19.9 | 25.4 | 103.6 | 140.6 | 206.6 | 607.4 | 362.7 | 551.9 | 306.4 | 159.5 | 59.3 |
| 3 | 210 | ASSAM & MEGHALAYA | 1904 | 11.1 | 56.1 | 51.9 | 457.1 | 375.2 | 385.7 | 477.6 | 438.8 | 245.9 | 115.9 | 46.4 |
| 4 | 211 | ASSAM & MEGHALAYA | 1905 | 19.9 | 16.9 | 137.9 | 213.0 | 275.5 | 521.7 | 439.1 | 649.1 | 276.0 | 200.0 | 16.8 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 110 | 317 | ASSAM & MEGHALAYA | 2011 | 11.1 | 11.4 | 109.0 | 92.1 | 238.3 | 316.0 | 395.8 | 302.6 | 221.6 | 30.2 | 11.9 |
| 111 | 318 | ASSAM & MEGHALAYA | 2012 | 15.2 | 6.9 | 28.8 | 279.1 | 185.8 | 729.7 | 444.3 | 289.2 | 411.6 | 199.4 | 17.1 |
| 112 | 319 | ASSAM & MEGHALAYA | 2013 | 1.1 | 9.6 | 44.0 | 112.8 | 346.7 | 286.2 | 367.8 | 289.7 | 229.3 | 126.3 | 1.0 |
| 113 | 320 | ASSAM & MEGHALAYA | 2014 | 2.0 | 28.3 | 29.3 | 51.5 | 351.1 | 426.4 | 374.4 | 484.6 | 420.2 | 35.0 | 3.0 |
| 114 | 321 | ASSAM & MEGHALAYA | 2015 | 13.4 | 15.5 | 37.5 | 250.9 | 332.5 | 558.5 | 300.1 | 590.9 | 279.9 | 62.6 | 14.0 |

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 | 207 | ASSAM & MEGHALAYA | 1901 | 27.1 | 19.5 | 30.6 | 223.0 | 207.0 | 524.9 | 430.6 | 464.1 | 291.4 | 163.7 | 115.6 |
| 1 | 208 | ASSAM & MEGHALAYA | 1902 | 9.3 | 10.2 | 105.6 | 350.0 | 262.1 | 620.7 | 510.8 | 536.0 | 441.3 | 97.0 | 7.8 |
| 2 | 209 | ASSAM & MEGHALAYA | 1903 | 19.9 | 25.4 | 103.6 | 140.6 | 206.6 | 607.4 | 362.7 | 551.9 | 306.4 | 159.5 | 59.3 |

| | index | SUBDIVISION | YEAR | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV |
|-----|-------|-------------------|------|------|------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| 3 | 210 | ASSAM & MEGHALAYA | 1904 | 11.1 | 56.1 | 51.9 | 457.1 | 375.2 | 385.7 | 477.6 | 438.8 | 245.9 | 115.9 | 46.4 |
| 4 | 211 | ASSAM & MEGHALAYA | 1905 | 19.9 | 16.9 | 137.9 | 213.0 | 275.5 | 521.7 | 439.1 | 649.1 | 276.0 | 200.0 | 16.8 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 110 | 317 | ASSAM & MEGHALAYA | 2011 | 11.1 | 11.4 | 109.0 | 92.1 | 238.3 | 316.0 | 395.8 | 302.6 | 221.6 | 30.2 | 11.9 |
| 111 | 318 | ASSAM & MEGHALAYA | 2012 | 15.2 | 6.9 | 28.8 | 279.1 | 185.8 | 729.7 | 444.3 | 289.2 | 411.6 | 199.4 | 17.1 |
| 112 | 319 | ASSAM & MEGHALAYA | 2013 | 1.1 | 9.6 | 44.0 | 112.8 | 346.7 | 286.2 | 367.8 | 289.7 | 229.3 | 126.3 | 1.0 |
| 113 | 320 | ASSAM & MEGHALAYA | 2014 | 2.0 | 28.3 | 29.3 | 51.5 | 351.1 | 426.4 | 374.4 | 484.6 | 420.2 | 35.0 | 3.0 |
| 114 | 321 | ASSAM & MEGHALAYA | 2015 | 13.4 | 15.5 | 37.5 | 250.9 | 332.5 | 558.5 | 300.1 | 590.9 | 279.9 | 62.6 | 14.0 |

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 | 264.000000 | 1958.000000 | 16.974783 | 31.441739 | 79.026957 | 203.115652 | 341.539130 | 510.161739 |
| std | 33.341666 | 33.341666 | 15.668601 | 20.963452 | 48.679075 | 74.338367 | 105.771005 | 106.248805 |
| min | 207.000000 | 1901.000000 | 0.100000 | 0.500000 | 4.400000 | 45.900000 | 159.700000 | 273.100000 |
| 25% | 235.500000 | 1929.500000 | 7.500000 | 15.800000 | 38.700000 | 151.350000 | 261.100000 | 421.550000 |
| 50% | 264.000000 | 1958.000000 | 12.800000 | 28.300000 | 75.700000 | 197.600000 | 325.700000 | 513.300000 |
| 75% | 292.500000 | 1986.500000 | 21.600000 | 41.950000 | 104.600000 | 251.500000 | 397.150000 | 583.650000 |
| max | 321.000000 | 2015.000000 | 83.900000 | 96.900000 | 253.800000 | 457.100000 | 681.200000 | 780.500000 |

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
```

```

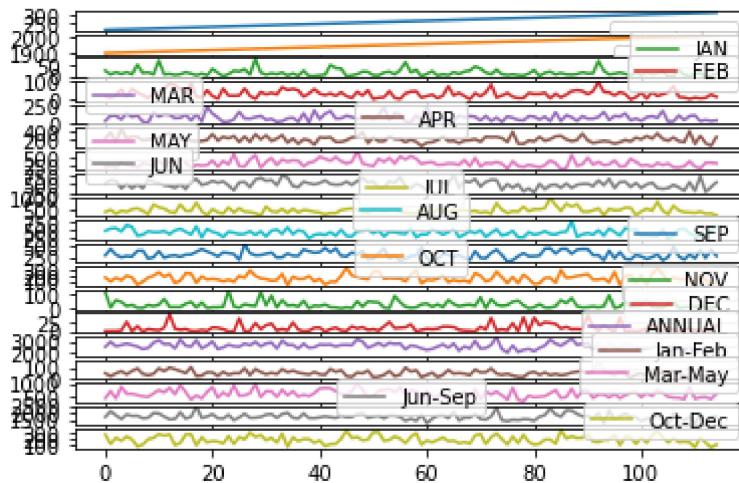
17 Mar-May      115 non-null    float64
18 Jun-Sep      115 non-null    float64
19 Oct-Dec      115 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 18.9+ KB

```

Line chart

In [11]: `df.plot.line(subplots=True)`

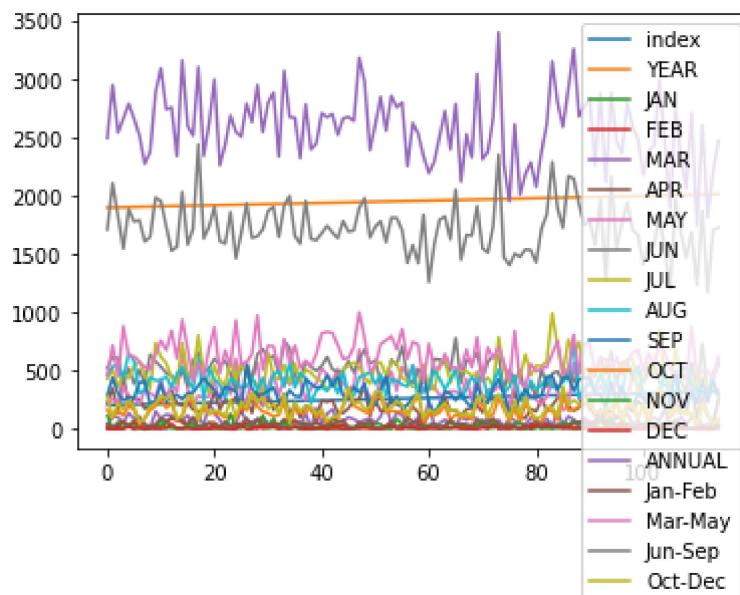
Out[11]: `array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>], dtype=object)`



Line chart

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

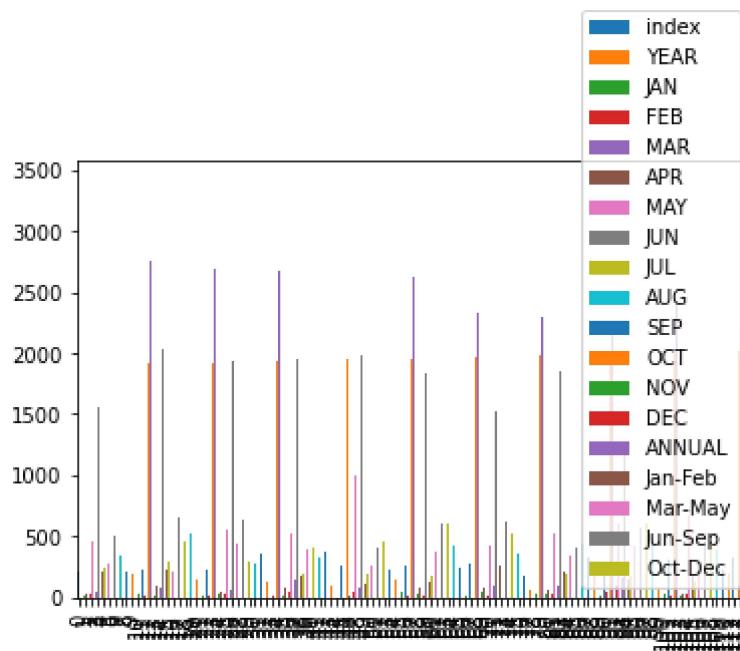
Out[12]: `<AxesSubplot:>`



Bar chart

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

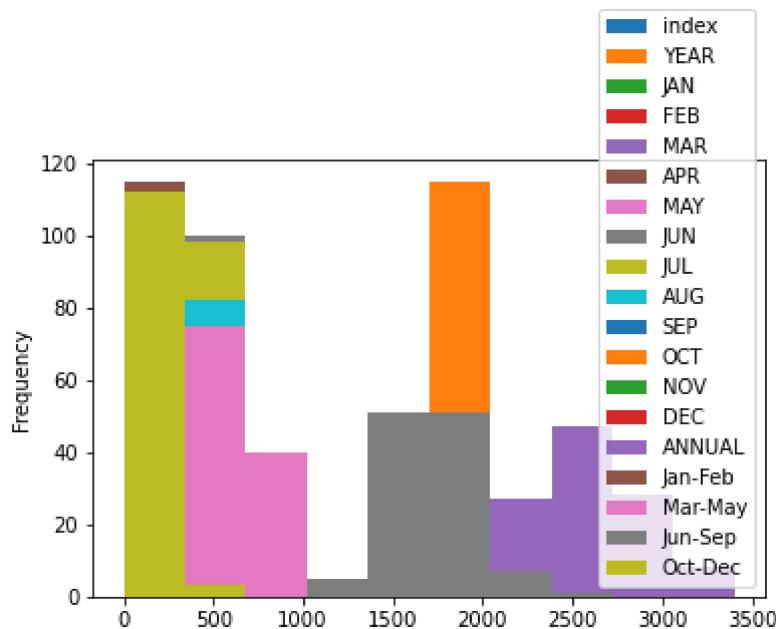
Out[13]: <AxesSubplot:>



Histogram

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

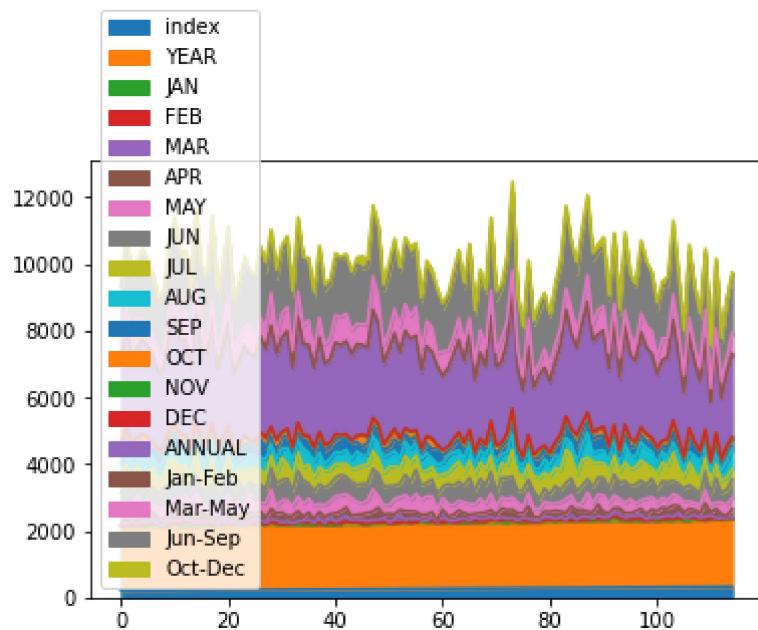
Out[14]: <AxesSubplot:ylabel='Frequency'>



Area chart

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

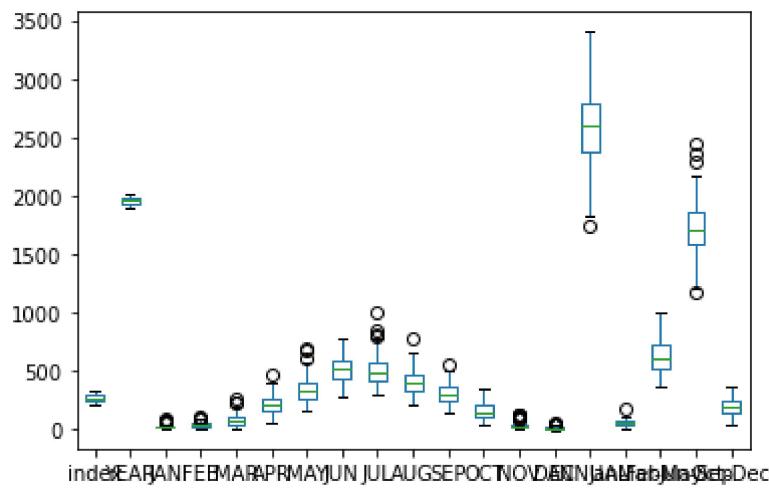
Out[15]: <AxesSubplot:>



Box chart

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

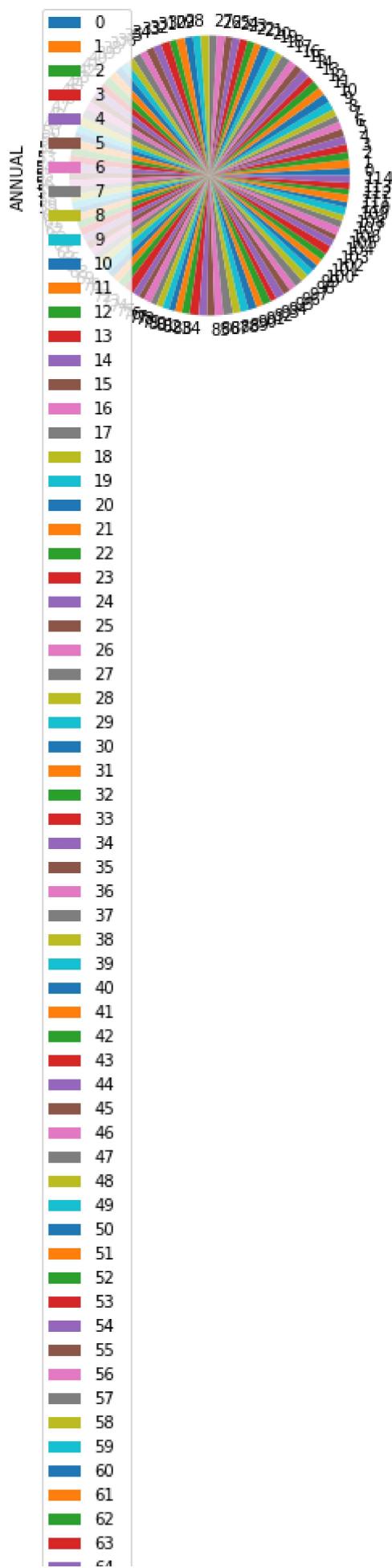
Out[16]: <AxesSubplot:>

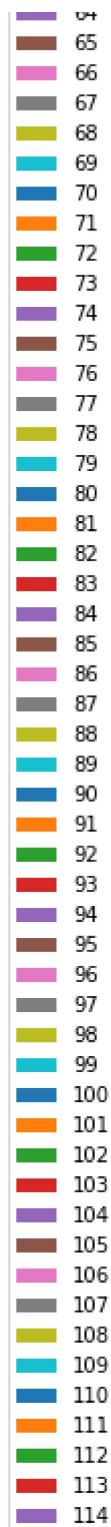


Pie chart

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

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

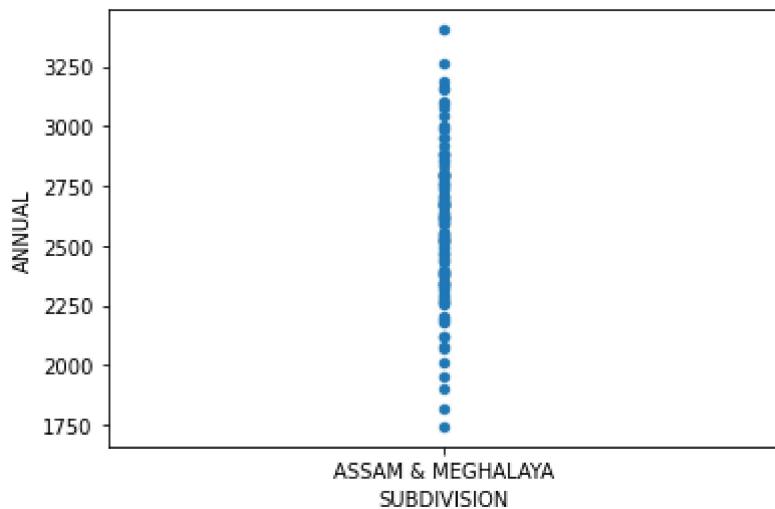




Scatter chart

```
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]:

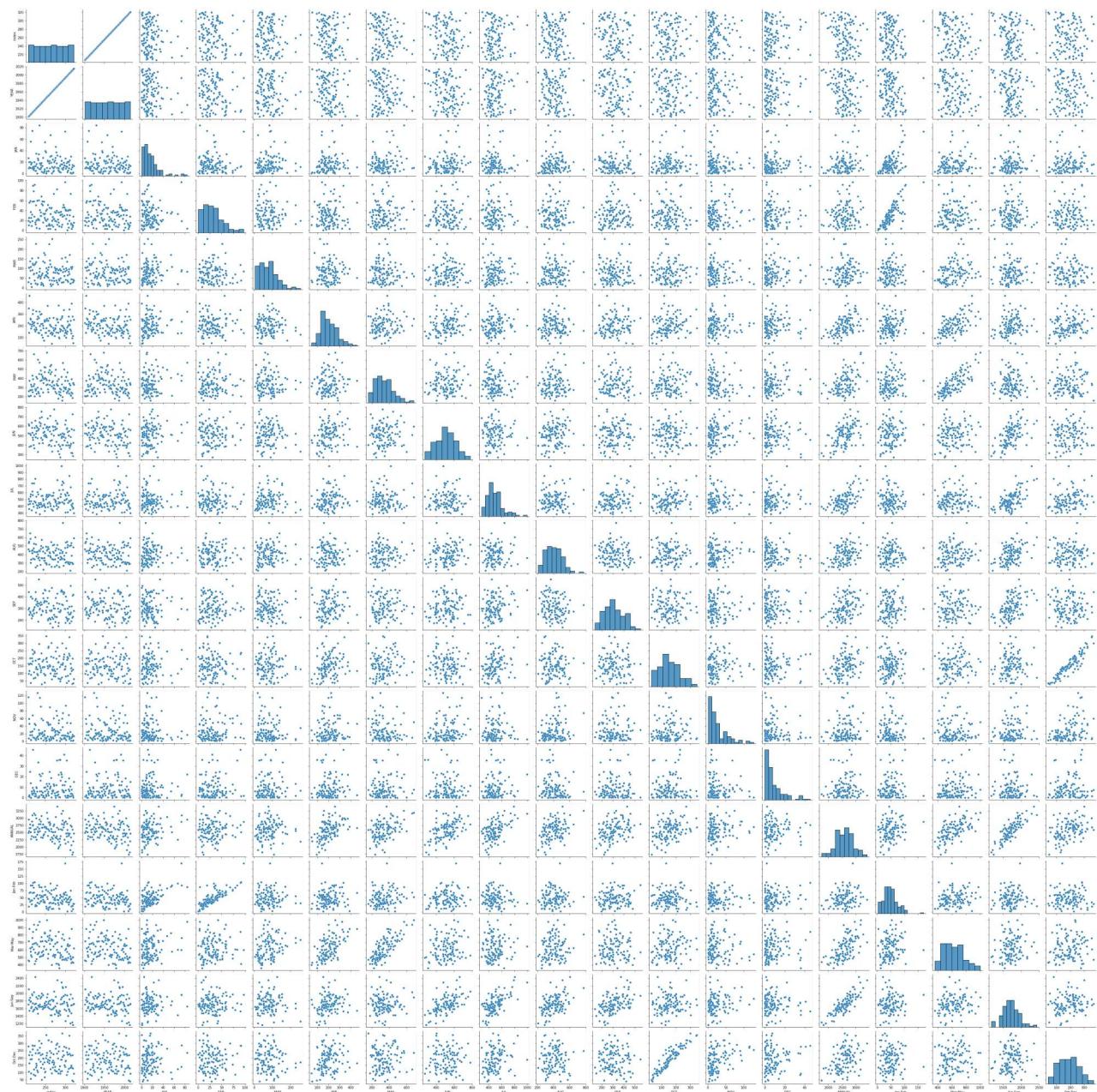
| | index | YEAR | JAN | FEB | MAR | APR | MAY | JUN |
|--------------|------------|-------------|------------|------------|------------|------------|------------|------------|
| count | 115.000000 | 115.000000 | 115.000000 | 115.000000 | 115.000000 | 115.000000 | 115.000000 | 115.000000 |
| mean | 264.000000 | 1958.000000 | 16.974783 | 31.441739 | 79.026957 | 203.115652 | 341.539130 | 510.161739 |
| std | 33.341666 | 33.341666 | 15.668601 | 20.963452 | 48.679075 | 74.338367 | 105.771005 | 106.248805 |
| min | 207.000000 | 1901.000000 | 0.100000 | 0.500000 | 4.400000 | 45.900000 | 159.700000 | 273.100000 |
| 25% | 235.500000 | 1929.500000 | 7.500000 | 15.800000 | 38.700000 | 151.350000 | 261.100000 | 421.550000 |

| | index | YEAR | JAN | FEB | MAR | APR | MAY | JUN |
|------------|------------|-------------|-----------|-----------|------------|------------|------------|------------|
| 50% | 264.000000 | 1958.000000 | 12.800000 | 28.300000 | 75.700000 | 197.600000 | 325.700000 | 513.300000 |
| 75% | 292.500000 | 1986.500000 | 21.600000 | 41.950000 | 104.600000 | 251.500000 | 397.150000 | 583.650000 |
| max | 321.000000 | 2015.000000 | 83.900000 | 96.900000 | 253.800000 | 457.100000 | 681.200000 | 780.500000 |

EDA AND VISUALIZATION

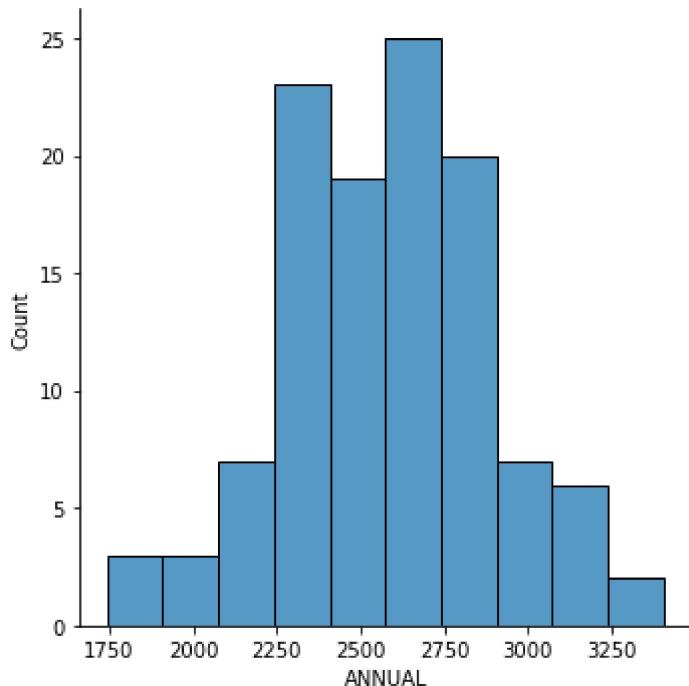
In [21]: `sns.pairplot(df)`

Out[21]: <seaborn.axisgrid.PairGrid at 0x220df242910>



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

Out[22]: <seaborn.axisgrid.FacetGrid at 0x220ea34e220>

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

Out[23]: <AxesSubplot:>

