# 20104016

# DEENA

# 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_south interior karnataka.csv")
df
```

Out[2]:

| | index | SUBDIVISION | YEAR | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 3772 | SOUTH INTERIOR KARNATAKA | 1901 | 4.9 | 31.8 | 3.0 | 32.7 | 109.6 | 106.0 | 210.0 | 109.2 | 140.8 | 170. |
| **1** | 3773 | SOUTH INTERIOR KARNATAKA | 1902 | 1.9 | 0.5 | 6.7 | 42.6 | 97.7 | 91.7 | 210.0 | 82.1 | 138.4 | 219. |
| **2** | 3774 | SOUTH INTERIOR KARNATAKA | 1903 | 0.3 | 0.0 | 1.1 | 11.6 | 125.1 | 129.7 | 284.4 | 155.7 | 197.1 | 154. |
| **3** | 3775 | SOUTH INTERIOR KARNATAKA | 1904 | 1.0 | 0.5 | 5.2 | 43.5 | 144.7 | 167.9 | 197.1 | 73.2 | 89.6 | 120. |
| **4** | 3776 | SOUTH INTERIOR KARNATAKA | 1905 | 1.7 | 7.9 | 14.2 | 23.6 | 118.6 | 95.9 | 148.4 | 140.6 | 43.1 | 142. |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **110** | 3882 | SOUTH INTERIOR KARNATAKA | 2011 | 2.1 | 12.4 | 12.4 | 80.2 | 83.5 | 177.1 | 202.4 | 199.5 | 111.2 | 144. |
| **111** | 3883 | SOUTH INTERIOR KARNATAKA | 2012 | 4.6 | 5.5 | 8.1 | 99.0 | 45.6 | 81.8 | 144.7 | 236.5 | 100.6 | 62. |
| **112** | 3884 | SOUTH INTERIOR KARNATAKA | 2013 | 0.5 | 10.1 | 11.7 | 34.6 | 95.6 | 176.2 | 307.4 | 151.7 | 191.8 | 103. |
| **113** | 3885 | SOUTH INTERIOR KARNATAKA | 2014 | 0.4 | 2.4 | 17.7 | 46.7 | 130.5 | 106.8 | 271.6 | 254.6 | 161.6 | 152. |
| **114** | 3886 | SOUTH INTERIOR KARNATAKA | 2015 | 1.7 | 0.2 | 24.4 | 80.5 | 125.3 | 218.7 | 112.0 | 136.6 | 164.5 | 106. |

115 rows × 20 columns

# Data Cleaning and Data Preprocessing

In [3]:
```
df=df.dropna()
```

In [4]:
```
df.columns
```

Out[4]:
```
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')
```

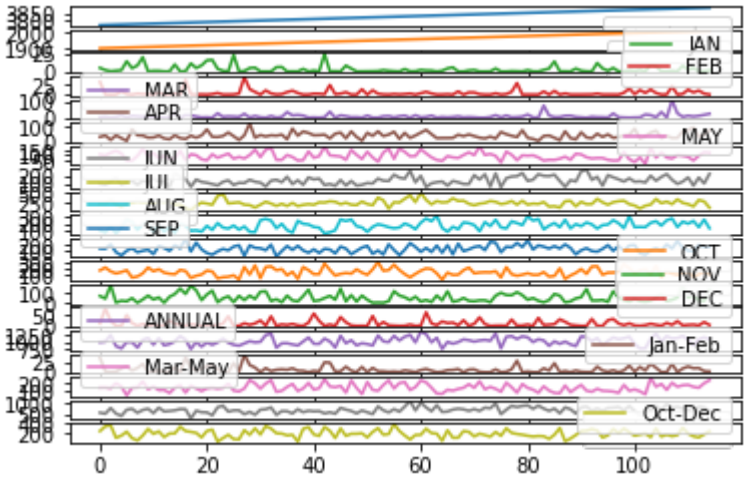
In [5]:
```
df.info()
```

```
<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 [6]:
```
df.plot.line(subplots=True)
```

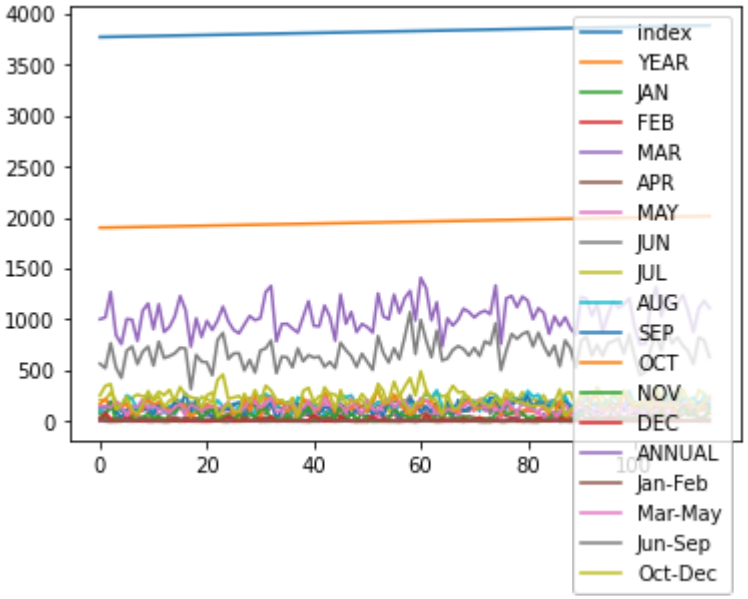
Out[6]:
```
array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>], dtype=object)
```




# Line chart

In [7]:
```
df.plot.line()
```

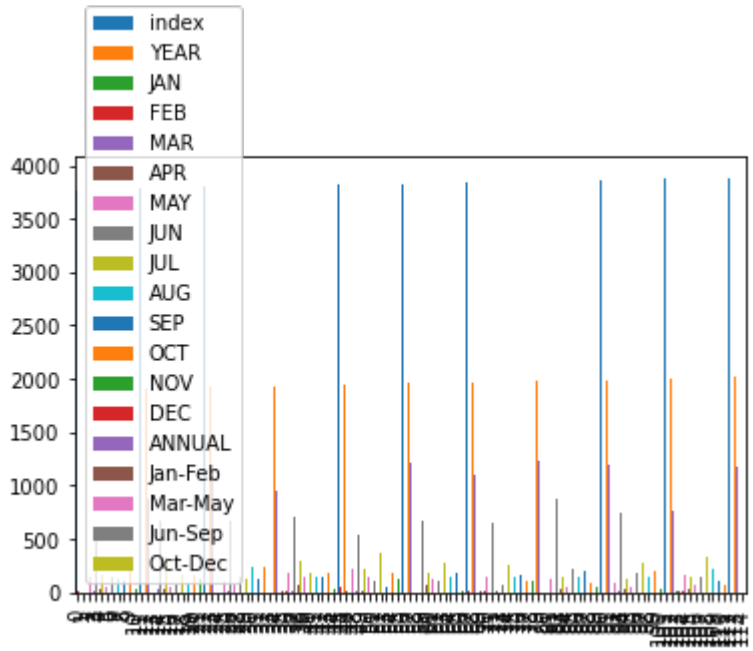
Out[7]:
```
<AxesSubplot:>
```




# Bar chart

In [8]:
```
df.plot.bar()
```

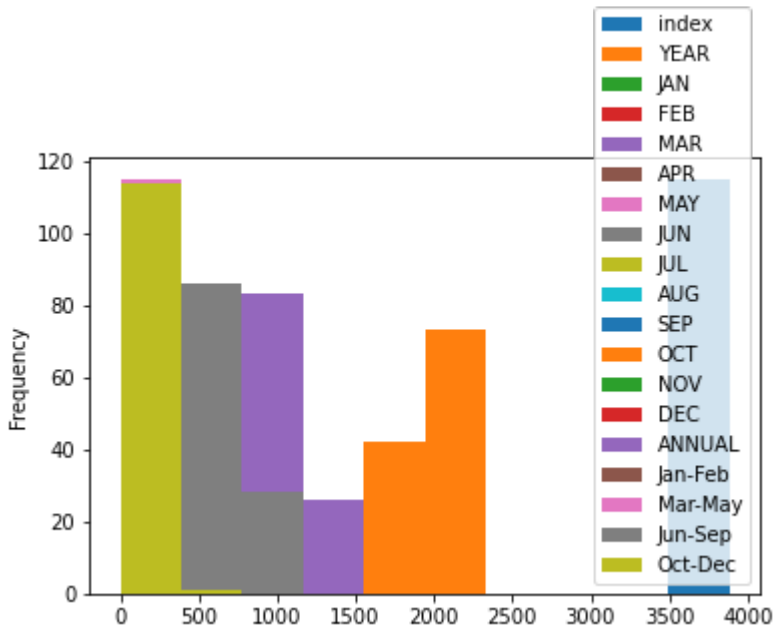
Out[8]: <AxesSubplot:>




# Histogram

In [9]:
```
df.plot.hist()
```

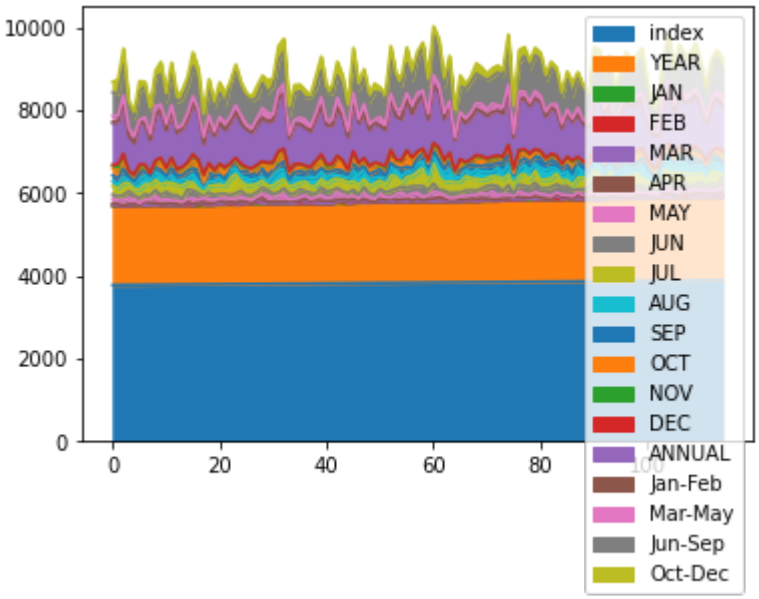
Out[9]: <AxesSubplot:ylabel='Frequency'>




# Area chart

In [10]:
```
df.plot.area()
```

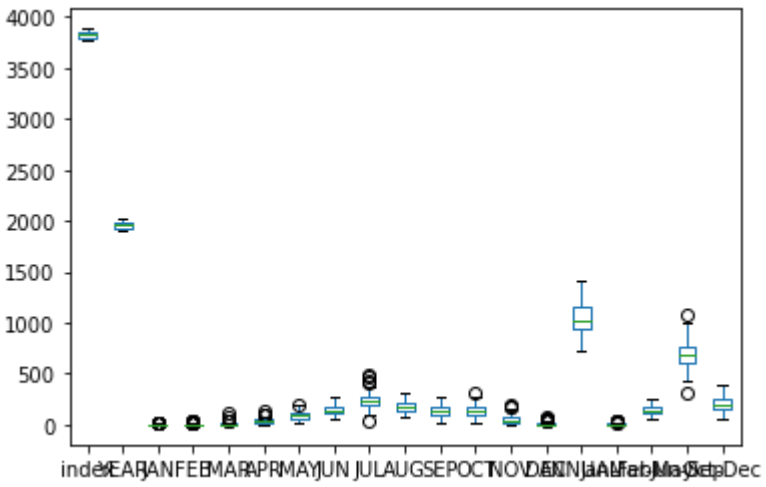
Out[10]: <AxesSubplot:>




# Box chart

In [11]:
```
df.plot.box()
```

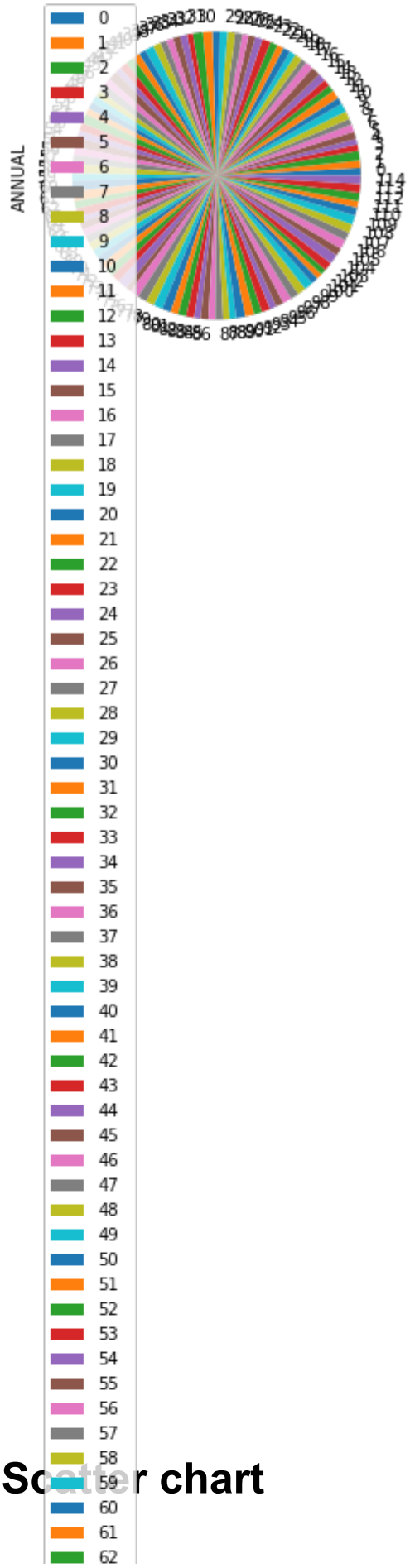
Out[11]: <AxesSubplot:>




# Pie chart

In [12]:
```
df.plot.pie(y='ANNUAL')
```

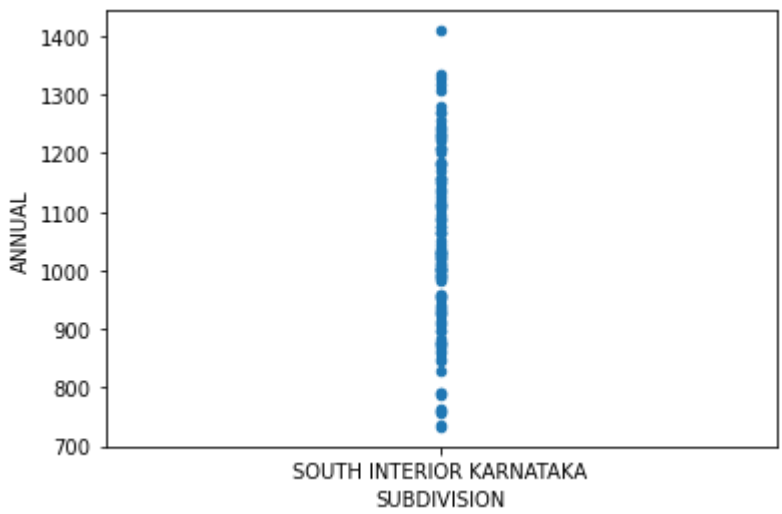
Out[12]:
```
<AxesSubplot:ylabel='ANNUAL'>
```




## Scatter chart

In [13]:
```
df.plot.scatter(x='SUBDIVISION', y='ANNUAL')
```

Out[13]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>




In [14]:
```
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
```

In [15]:
```
df.describe()
```

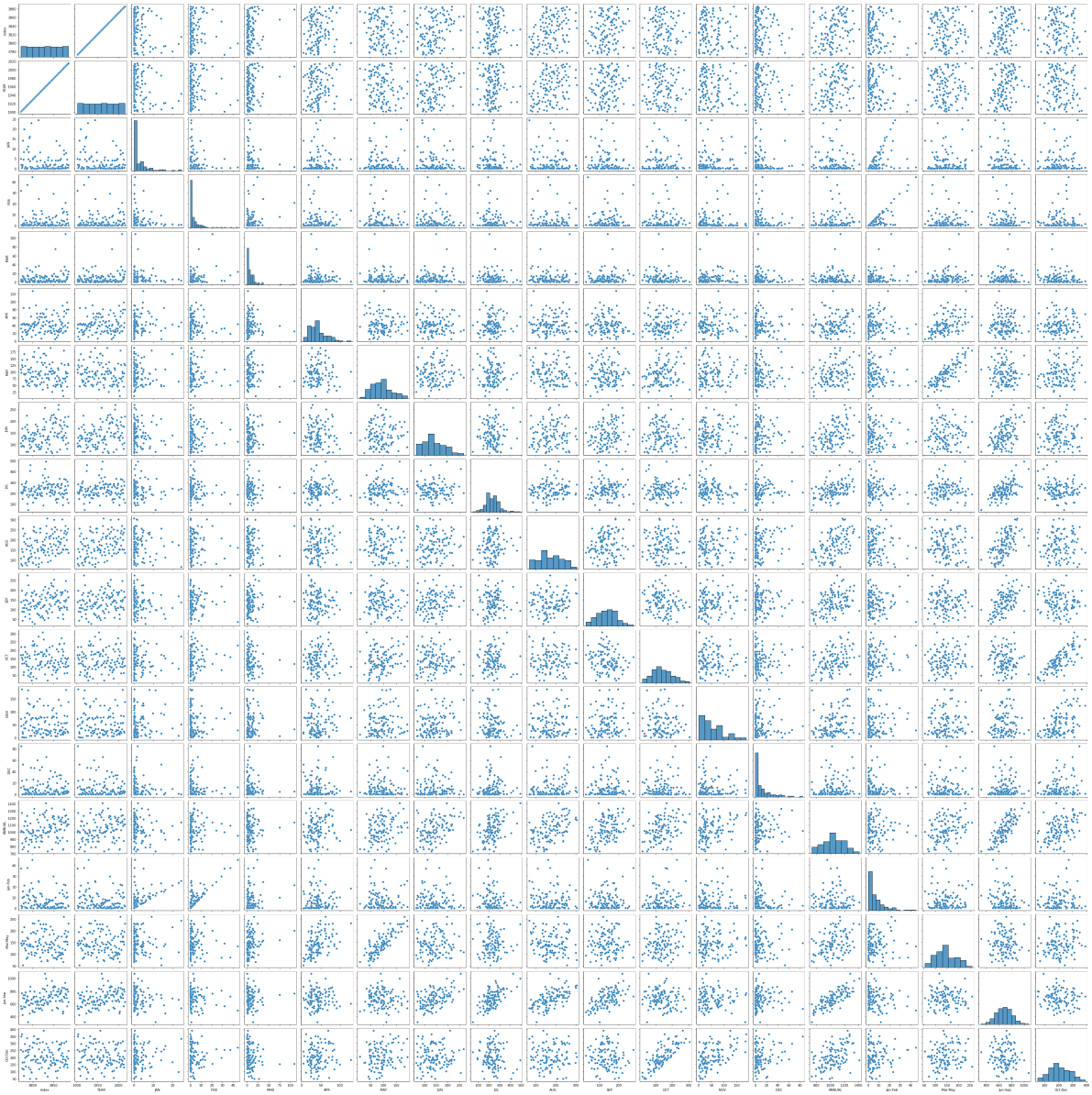
Out[15]:

| | index | YEAR | JAN | FEB | MAR | APR | MAY | |
|---|---|---|---|---|---|---|---|---|
| **count** | 115.000000 | 115.000000 | 115.000000 | 115.000000 | 115.000000 | 115.000000 | 115.000000 | 11 |
| **mean** | 3829.000000 | 1958.000000 | 2.928696 | 4.163478 | 9.485217 | 42.280870 | 92.100000 | 14 |
| **std** | 33.341666 | 33.341666 | 4.808741 | 7.560439 | 13.955998 | 22.061039 | 38.219243 | 4 |
| **min** | 3772.000000 | 1901.000000 | 0.000000 | 0.000000 | 0.000000 | 4.800000 | 9.600000 | 6 |
| **25%** | 3800.500000 | 1929.500000 | 0.100000 | 0.100000 | 2.300000 | 25.450000 | 63.350000 | 10 |
| **50%** | 3829.000000 | 1958.000000 | 0.800000 | 1.200000 | 5.200000 | 40.000000 | 90.900000 | 13 |
| **75%** | 3857.500000 | 1986.500000 | 4.100000 | 4.600000 | 12.750000 | 51.400000 | 114.150000 | 17 |
| **max** | 3886.000000 | 2015.000000 | 24.400000 | 44.300000 | 108.900000 | 127.700000 | 190.500000 | 26 |

# EDA AND VISUALIZATION

In [16]:
```
sns.pairplot(df)
```

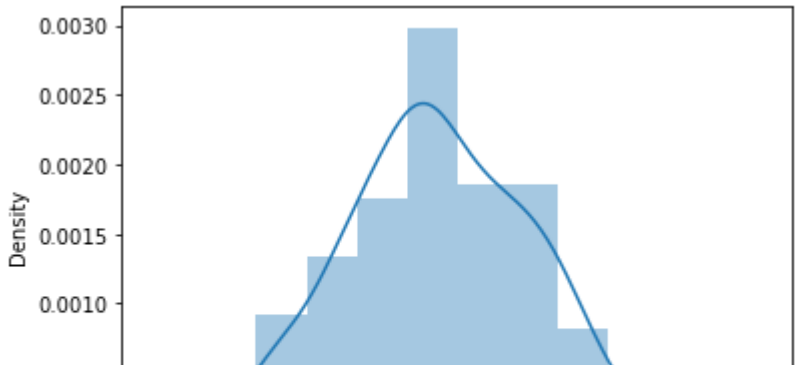
Out[16]:
```
<seaborn.axisgrid.PairGrid at 0x1e4db437100>
```



In [17]:
```
sns.distplot(df['ANNUAL'])
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: Fut
ureWarning: `distplot` is a deprecated function and will be removed in a futu
re version. Please adapt your code to use either `displot` (a figure-level fu
nction with similar flexibility) or `histplot` (an axes-level function for hi
stograms).
  warnings.warn(msg, FutureWarning)
```

Out[17]: <AxesSubplot:xlabel='ANNUAL', ylabel='Density'>




In [18]:
```
sns.heatmap(df.corr())
```

Out[18]: <AxesSubplot:>