Type *Markdown* and LaTeX: $\alpha^2$

# 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(r"C:\Users\user\Downloads\drive-download-20230804T043023Z-001\rainfall_orissa.csv"
df
```

Out[2]:

| | index | SUBDIVISION | YEAR | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC | ANNUAL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 2622 | MADHYA MAHARASHTRA | 1901 | 18.8 | 0.6 | 7.7 | 36.6 | 30.4 | 107.7 | 215.9 | 194.1 | 83.7 | 68.7 | 4.4 | 0.5 | 769. |
| **1** | 2623 | MADHYA MAHARASHTRA | 1902 | 7.8 | 0.0 | 0.1 | 5.0 | 9.8 | 102.6 | 210.9 | 114.5 | 169.5 | 60.4 | 40.5 | 62.9 | 784. |
| **2** | 2624 | MADHYA MAHARASHTRA | 1903 | 7.6 | 0.0 | 0.0 | 3.2 | 77.2 | 86.3 | 281.8 | 155.5 | 142.3 | 74.2 | 7.6 | 2.2 | 837. |
| **3** | 2625 | MADHYA MAHARASHTRA | 1904 | 0.4 | 4.7 | 1.7 | 3.0 | 18.7 | 114.6 | 126.5 | 59.5 | 183.0 | 91.1 | 0.0 | 0.4 | 603. |
| **4** | 2626 | MADHYA MAHARASHTRA | 1905 | 0.0 | 1.2 | 0.0 | 2.3 | 23.6 | 65.0 | 252.8 | 79.0 | 52.6 | 52.9 | 8.3 | 0.0 | 537. |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **110** | 2732 | MADHYA MAHARASHTRA | 2011 | 0.0 | 0.3 | 0.3 | 5.0 | 2.9 | 133.3 | 261.4 | 238.1 | 148.4 | 62.8 | 0.0 | 0.0 | 852. |
| **111** | 2733 | MADHYA MAHARASHTRA | 2012 | 0.0 | 0.0 | 0.0 | 3.0 | 1.4 | 67.9 | 203.0 | 187.8 | 129.5 | 95.2 | 2.2 | 0.0 | 689. |
| **112** | 2734 | MADHYA MAHARASHTRA | 2013 | 0.1 | 5.3 | 0.8 | 5.7 | 6.0 | 212.4 | 311.8 | 147.0 | 210.3 | 57.8 | 4.0 | 1.3 | 962. |
| **113** | 2735 | MADHYA MAHARASHTRA | 2014 | 3.1 | 6.2 | 24.4 | 7.5 | 29.8 | 44.0 | 277.9 | 240.3 | 120.4 | 38.5 | 32.8 | 13.1 | 838. |
| **114** | 2736 | MADHYA MAHARASHTRA | 2015 | 1.4 | 0.8 | 41.2 | 9.6 | 24.4 | 177.0 | 111.7 | 67.2 | 146.6 | 48.3 | 16.2 | 0.1 | 644. |

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')
```

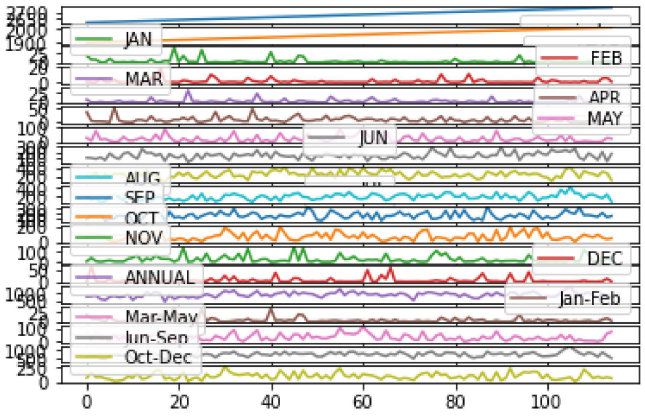
In [5]:
```
df.info()
```

```
<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)
```

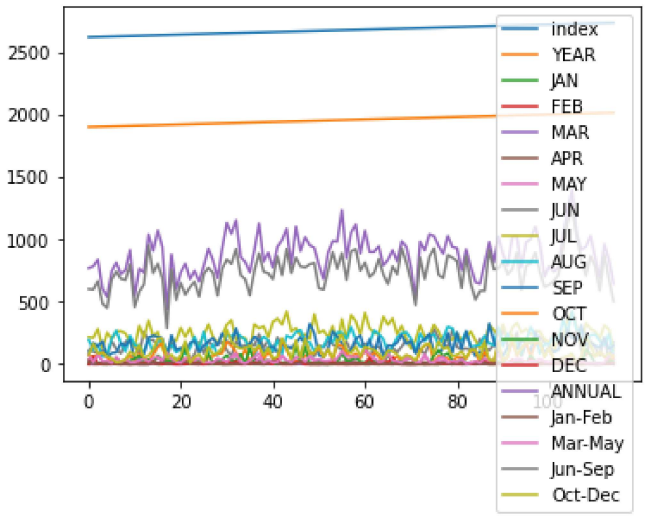
Out[6]:
```
array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
    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




# Line chart

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

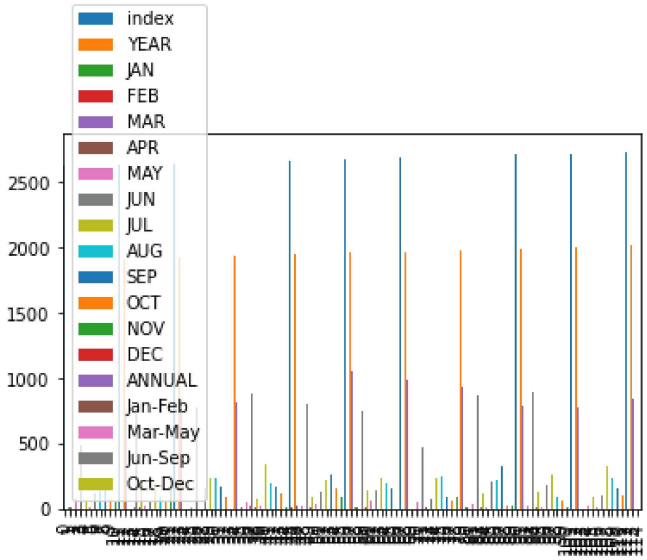
Out[7]: <AxesSubplot:>




# Bar chart

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

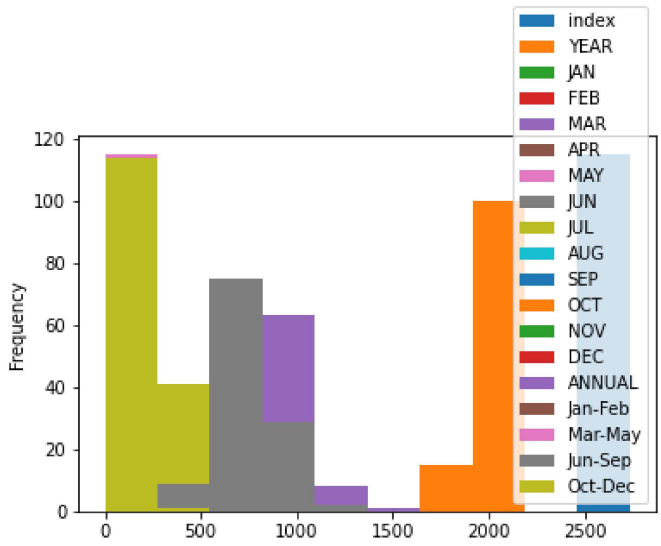
Out[8]: <AxesSubplot:>




# Histogram

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

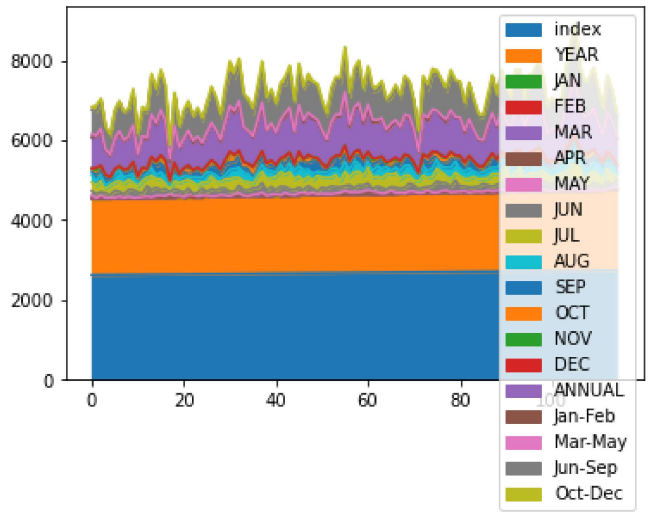
Out[9]: <AxesSubplot:ylabel='Frequency'>




# Area chart

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

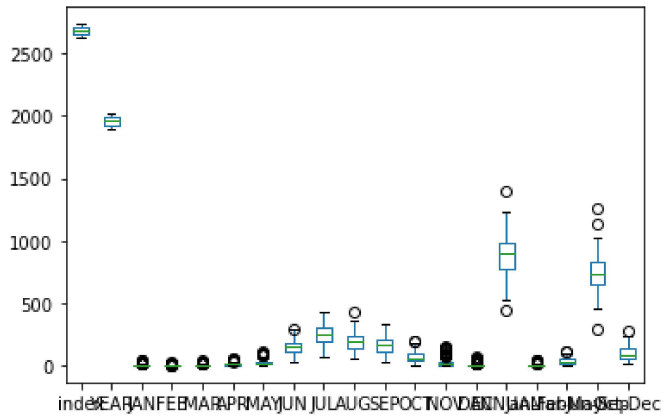
Out[10]: <AxesSubplot:>




# Box chart

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

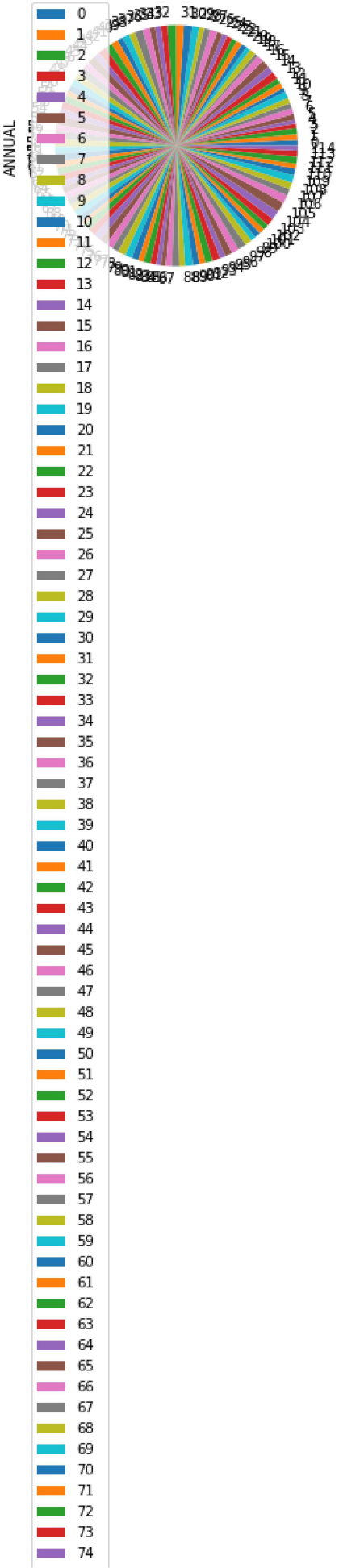
Out[11]: <AxesSubplot:>

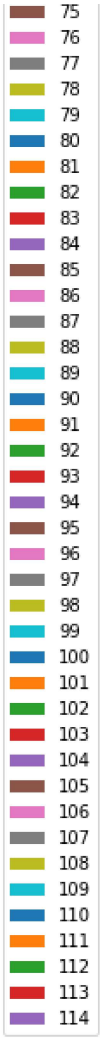



# Pie chart

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

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

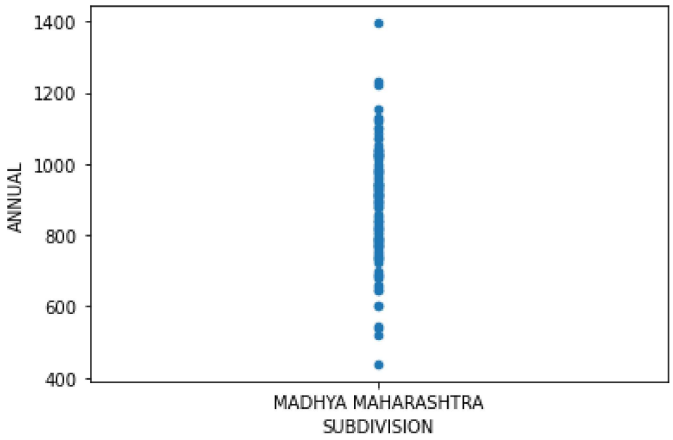

ANNUAL
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74




# Scatter chart

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

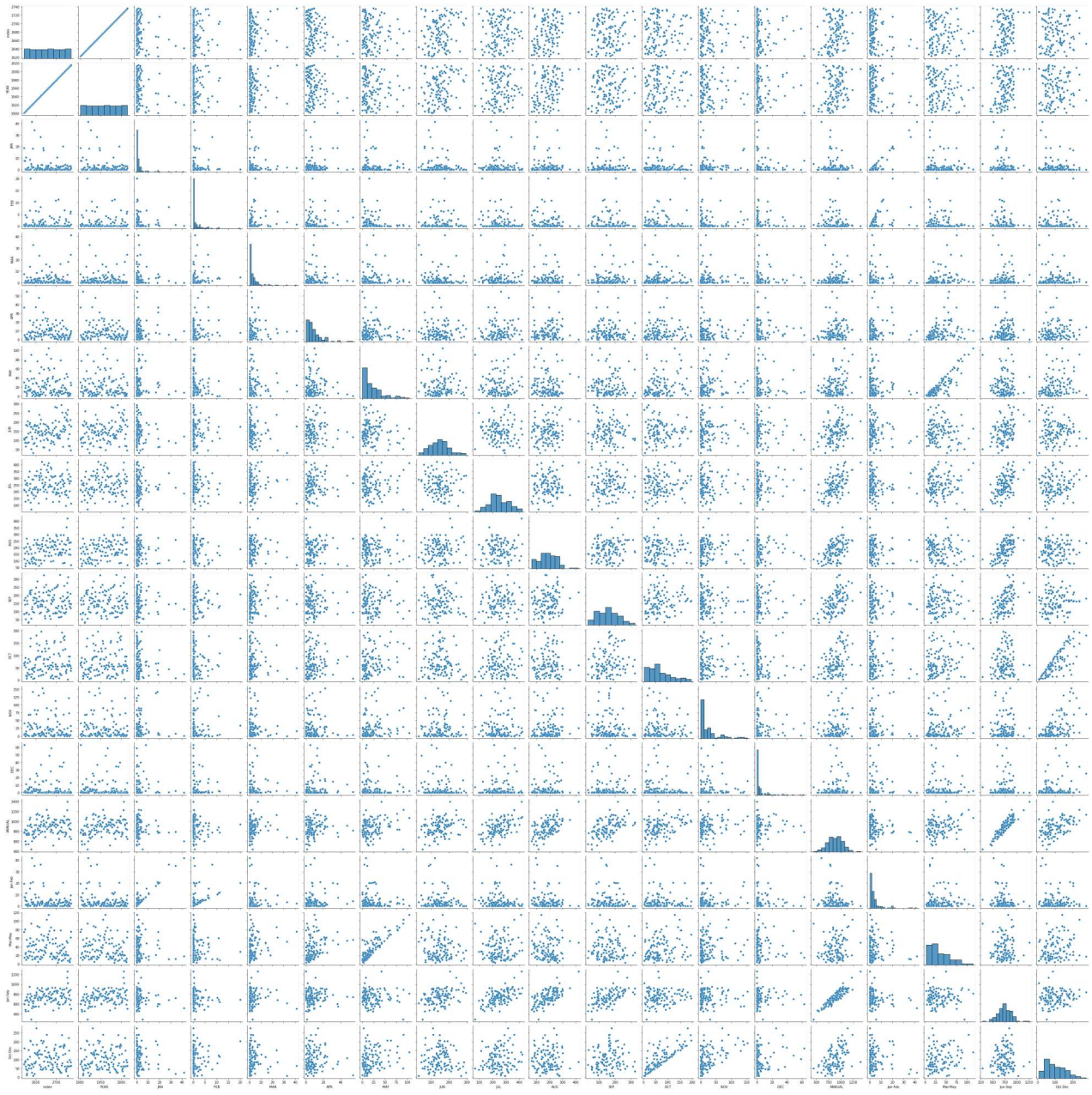
Out[13]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>




# Seaborn

In [14]:
```
sns.pairplot(df)
```

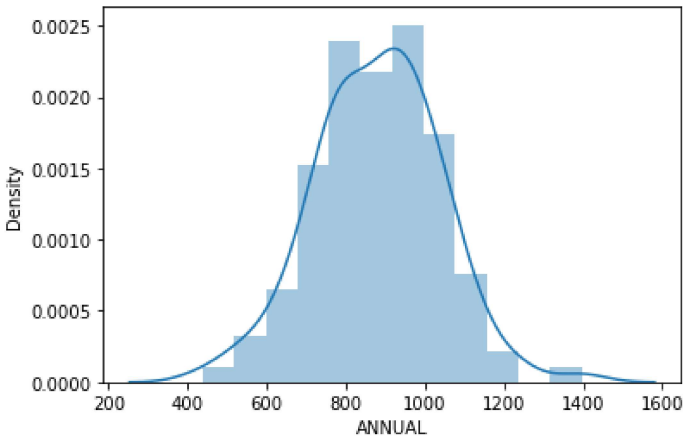
Out[14]:
```
<seaborn.axisgrid.PairGrid at 0x22ea4bfde80>
```



In [15]:
```
sns.distplot(df['ANNUAL'])
```

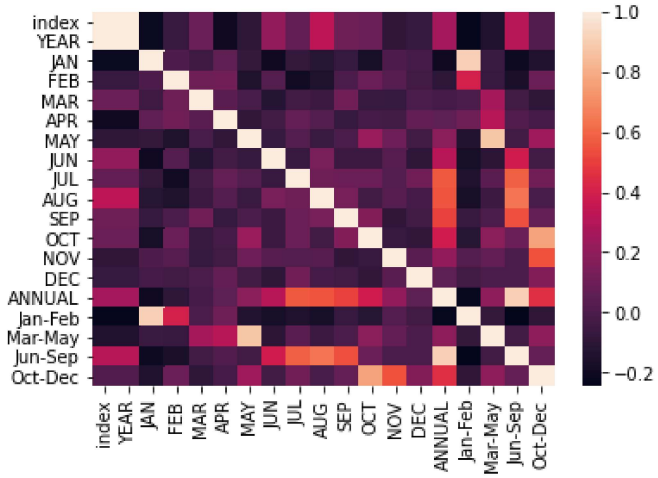
```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distp
lot` is a deprecated function and will be removed in a future version. Please adapt your code t
o use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axe
s-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[15]: <AxesSubplot:xlabel='ANNUAL', ylabel='Density'>




In [16]:
```
sns.heatmap(df.corr())
```

Out[16]: <AxesSubplot:>




In [ ]: