Importing Libraries ¶

In [4]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

Importing Datasets

In [5]:
```
df=pd.read_csv("rainfall_arunachal pradesh.csv")
df
```

Out[5]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	O
0	110	ARUNACHAL PRADESH	1916	48.1	69.8	71.1	316.1	424.6	1124.9	NaN	629.7	333.9	N
1	111	ARUNACHAL PRADESH	1917	21.4	164.5	NaN	269.6	107.9	823.8	909.1	628.4	411.5	19
2	112	ARUNACHAL PRADESH	1918	10.4	11.0	191.2	144.6	861.1	1609.9	1303.0	692.6	515.8	12
3	113	ARUNACHAL PRADESH	1919	34.5	67.8	28.5	256.9	420.6	973.6	999.0	286.7	628.7	94
4	114	ARUNACHAL PRADESH	1920	14.0	196.3	605.6	364.7	173.6	840.6	535.4	896.5	376.7	10
...	...	...	...	...	...	...	...	...	...	...	...	...	...
92	202	ARUNACHAL PRADESH	2011	40.0	51.3	174.5	240.8	219.6	288.4	531.4	277.6	286.7	5
93	203	ARUNACHAL PRADESH	2012	57.8	35.8	134.2	403.4	187.4	645.8	638.9	316.0	724.9	24
94	204	ARUNACHAL PRADESH	2013	18.5	40.5	115.1	175.1	335.8	290.0	329.6	230.2	316.1	16
95	205	ARUNACHAL PRADESH	2014	19.0	101.9	80.3	86.7	299.0	415.8	392.4	599.6	343.0	3
96	206	ARUNACHAL PRADESH	2015	30.8	47.5	97.5	287.1	238.9	637.9	329.3	595.5	374.2	6

97 rows × 20 columns

Data Cleaning and Data Preprocessing

In [6]:
```
df=df.dropna()
```

In [7]:
```
df.columns
```

Out[7]:
```
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')
```

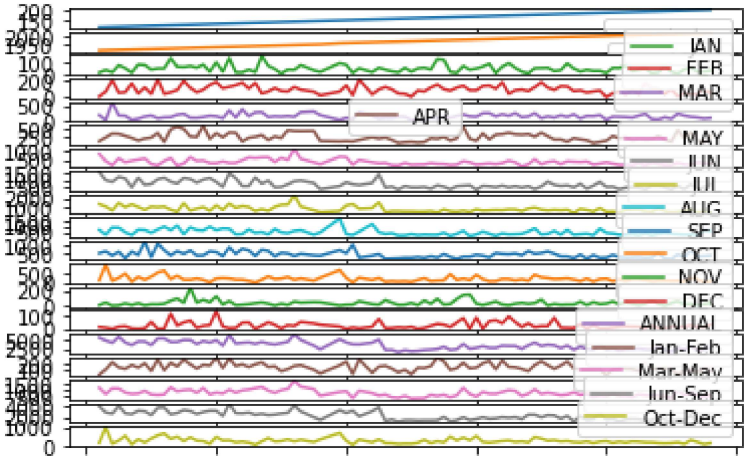
In [8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 91 entries, 2 to 96
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   index        91 non-null     int64  
 1   SUBDIVISION  91 non-null     object 
 2   YEAR         91 non-null     int64  
 3   JAN          91 non-null     float64
 4   FEB          91 non-null     float64
 5   MAR          91 non-null     float64
 6   APR          91 non-null     float64
 7   MAY          91 non-null     float64
 8   JUN          91 non-null     float64
 9   JUL          91 non-null     float64
 10  AUG          91 non-null     float64
 11  SEP          91 non-null     float64
 12  OCT          91 non-null     float64
 13  NOV          91 non-null     float64
 14  DEC          91 non-null     float64
 15  ANNUAL       91 non-null     float64
 16  Jan-Feb      91 non-null     float64
 17  Mar-May      91 non-null     float64
 18  Jun-Sep      91 non-null     float64
 19  Oct-Dec      91 non-null     float64
dtypes: float64(17), int64(2), object(1)
memory usage: 14.9+ KB
```

Line chart

In [9]:
```
df.plot.line(subplots=True)
```

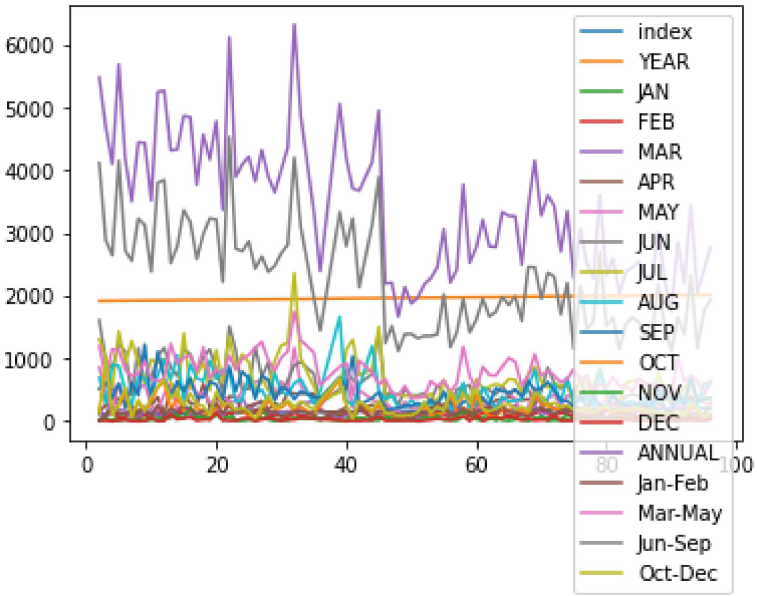
Out[9]:
```
array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>], dtype=object)
```



Line chart

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

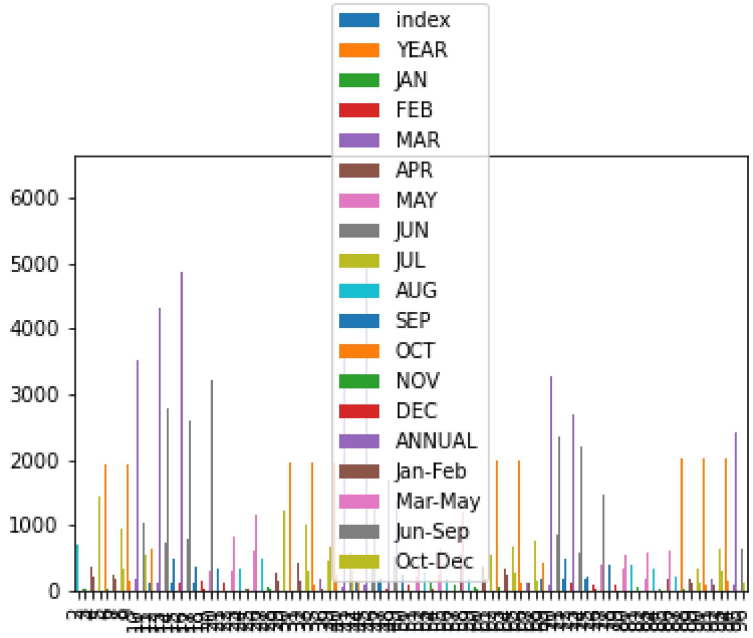
Out[10]:
```
<AxesSubplot:>
```



Bar chart

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

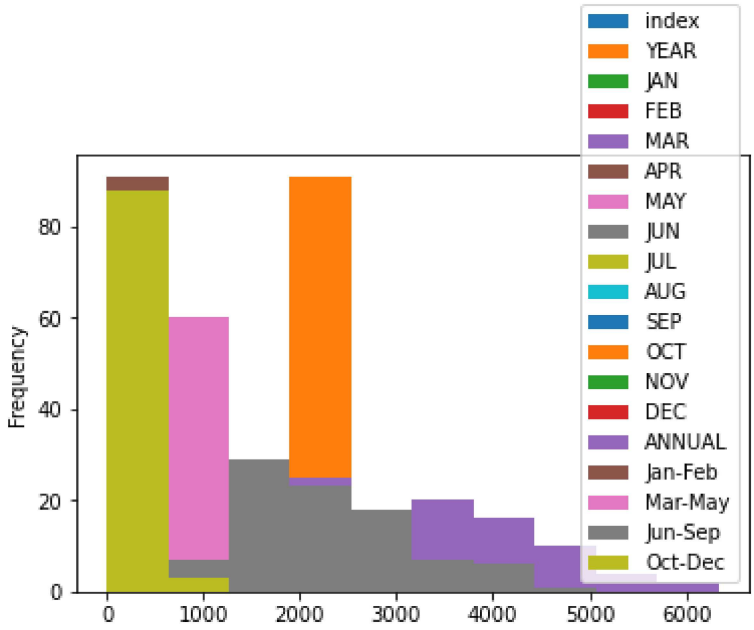
Out[11]: <AxesSubplot:>



Histogram

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

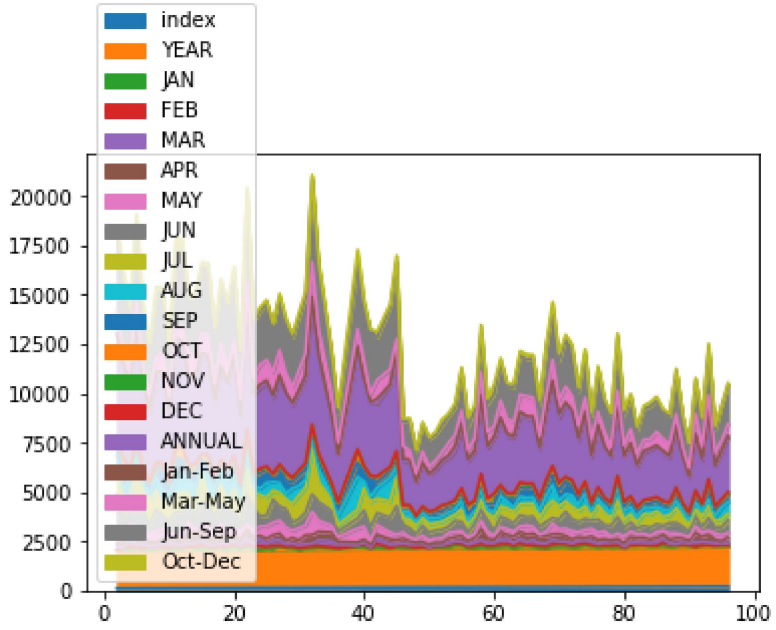
Out[12]: <AxesSubplot:ylabel='Frequency'>



Area chart

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

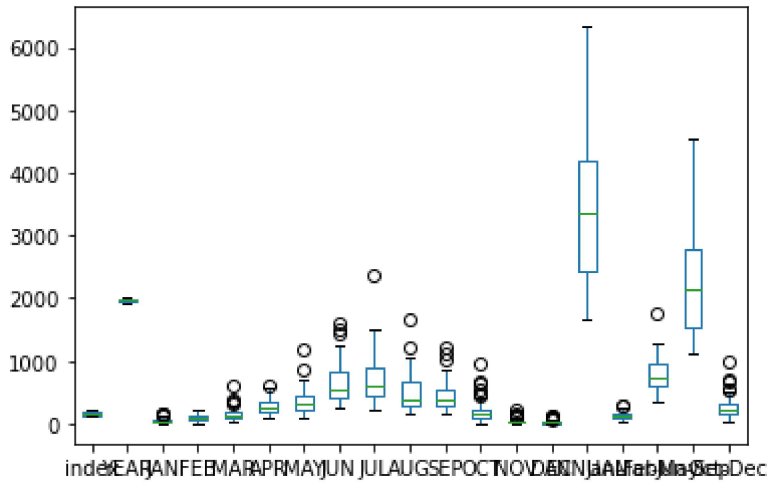
Out[13]: <AxesSubplot:>



Box chart

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

Out[14]: <AxesSubplot:>

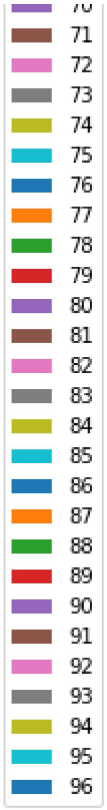


Pie chart

In [15]:
```
df.plot.pie(y='ANNUAL' )
```

Out[15]:
```
<AxesSubplot:ylabel='ANNUAL'>
```

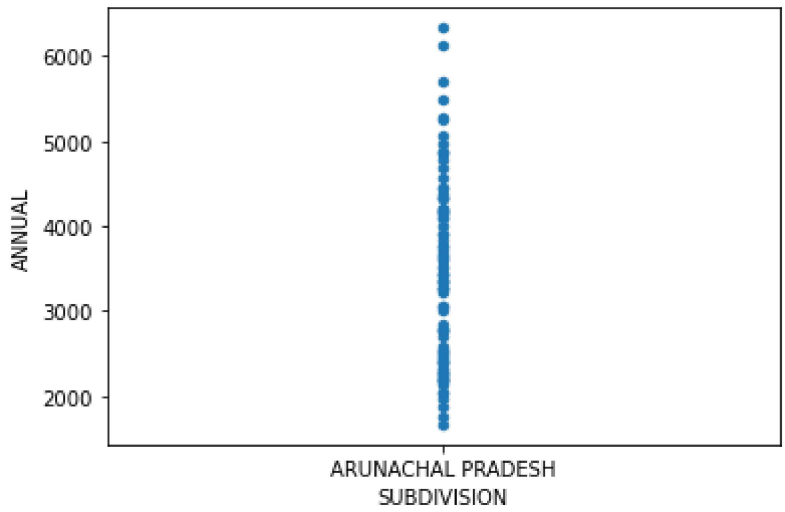

ANNUAL
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
35
36
39
40
41
42
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69



Scatter chart

In [16]:
```
df.plot.scatter(x='SUBDIVISION' ,y='ANNUAL')
```

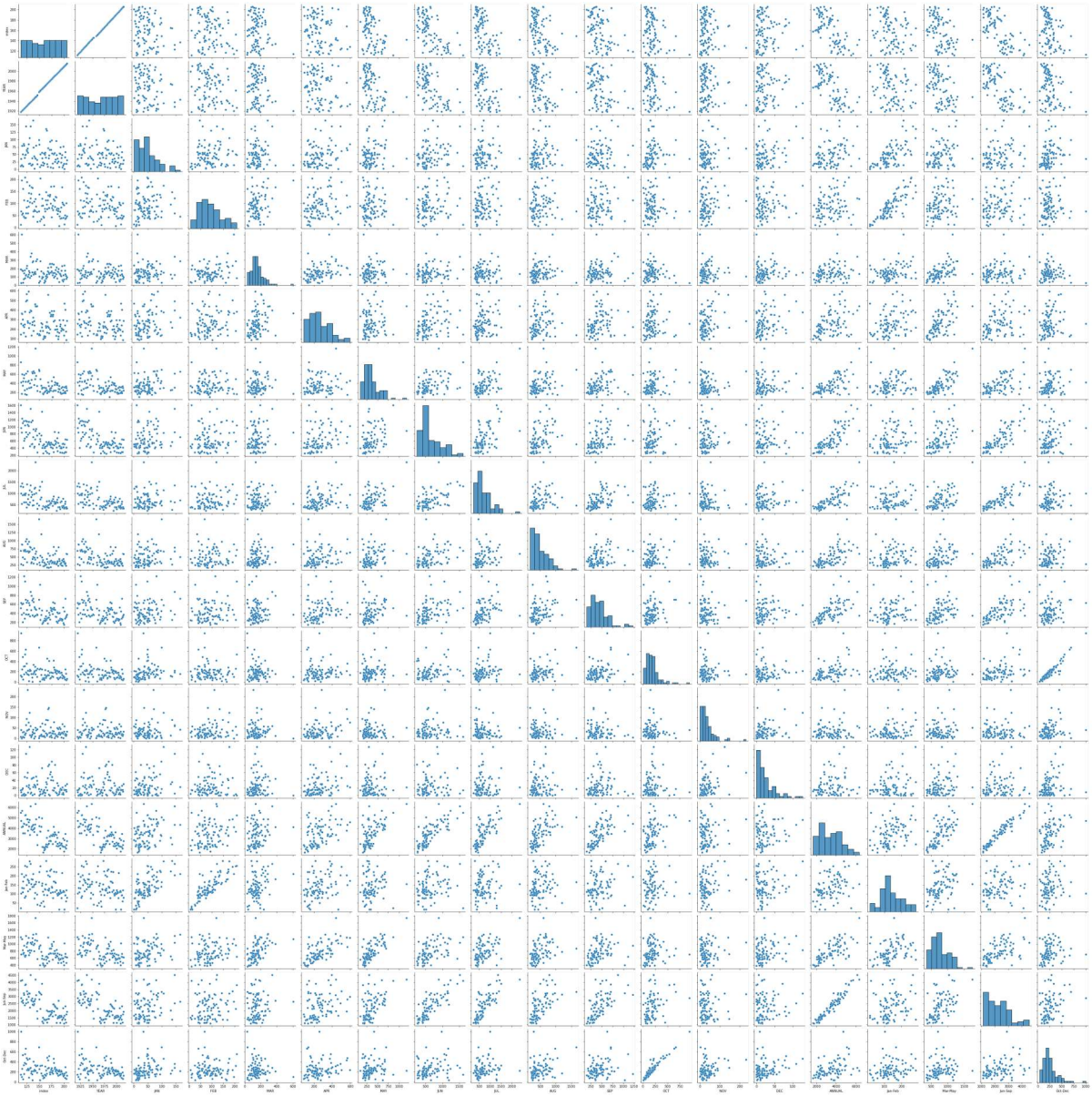
Out[16]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>



Seaborn

In [17]:
```
sns.pairplot(df)
```

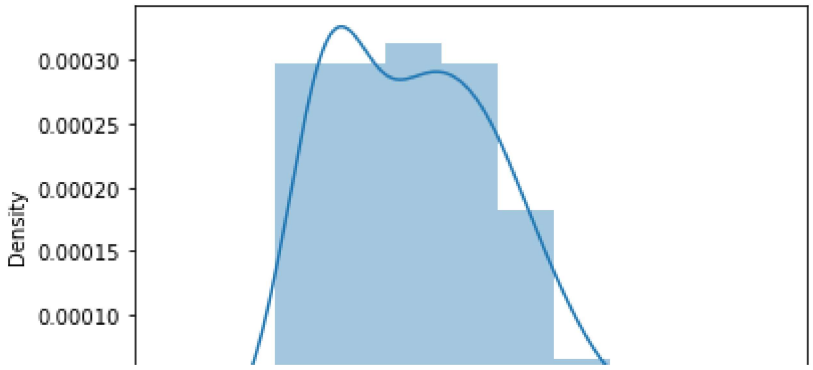
Out[17]:
```
<seaborn.axisgrid.PairGrid at 0x1f179cb27f0>
```



In [18]:
```
sns.distplot(df['ANNUAL'])
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: F
utureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-le
vel function with similar flexibility) or `histplot` (an axes-level functio
n for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[18]: <AxesSubplot:xlabel='ANNUAL', ylabel='Density'>



In [19]:
```
sns.heatmap(df.corr())
```

Out[19]: <AxesSubplot:>