Type *Markdown* and LaTeX: α^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_bihar.csv"
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUA
0	897	BIHAR	1901	51.8	19.6	11.9	1.1	65.6	66.3	245.9	319.4	155.1	8.3	7.3	0.1	952
1	898	BIHAR	1902	4.6	0.7	24.3	17.3	66.3	118.2	361.0	225.5	358.7	28.5	1.1	0.0	1206
2	899	BIHAR	1903	5.3	4.7	2.0	4.7	28.2	192.9	115.0	342.6	173.9	147.0	0.1	0.0	1016
3	900	BIHAR	1904	6.3	1.7	3.5	5.3	118.7	191.6	394.4	351.3	84.4	98.1	10.6	3.8	1269
4	901	BIHAR	1905	16.0	30.1	32.6	21.4	77.5	50.5	409.1	495.3	353.9	11.6	0.0	0.6	1498
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
110	1007	BIHAR	2011	4.2	7.7	9.2	23.9	74.5	211.0	241.1	278.7	234.1	10.0	2.0	0.9	1097
111	1008	BIHAR	2012	18.1	2.7	7.3	20.4	18.8	96.2	354.0	240.4	233.8	34.3	6.4	0.0	1032
112	1009	BIHAR	2013	5.1	22.6	0.6	32.3	89.5	183.3	182.0	213.6	143.3	197.1	0.4	0.0	1069
113	1010	BIHAR	2014	17.0	33.5	8.4	0.7	103.9	115.2	265.4	307.6	160.3	47.8	0.0	1.2	1061
114	1011	BIHAR	2015	12.8	1.8	27.2	38.7	39.5	122.1	231.5	287.0	101.7	10.4	0.0	0.0	872

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')
```

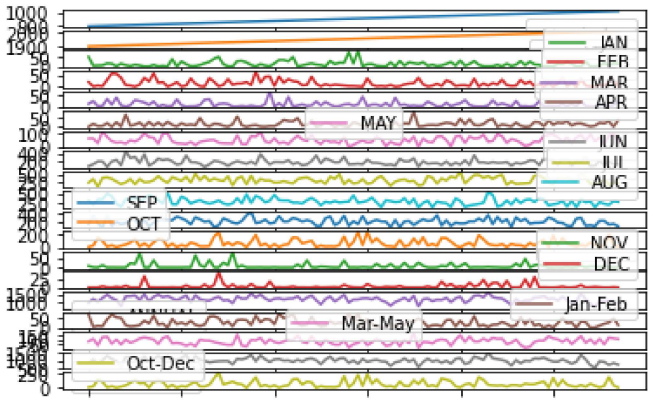
In [5]:
```
df.info()
```

```
<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)
```

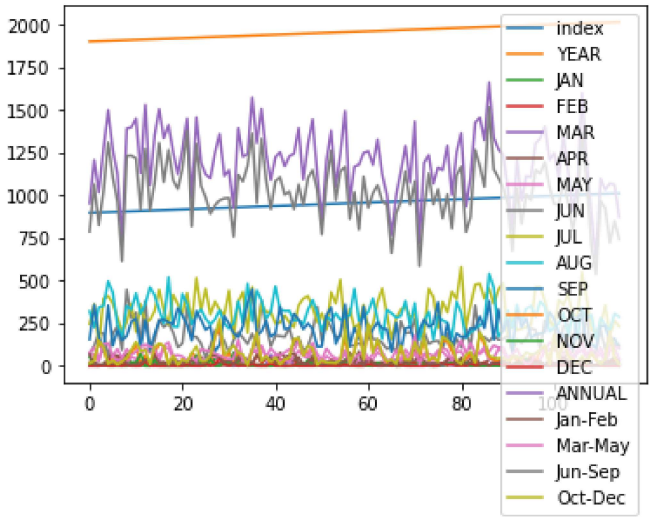
Out[6]:
```
array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
    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



Line chart

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

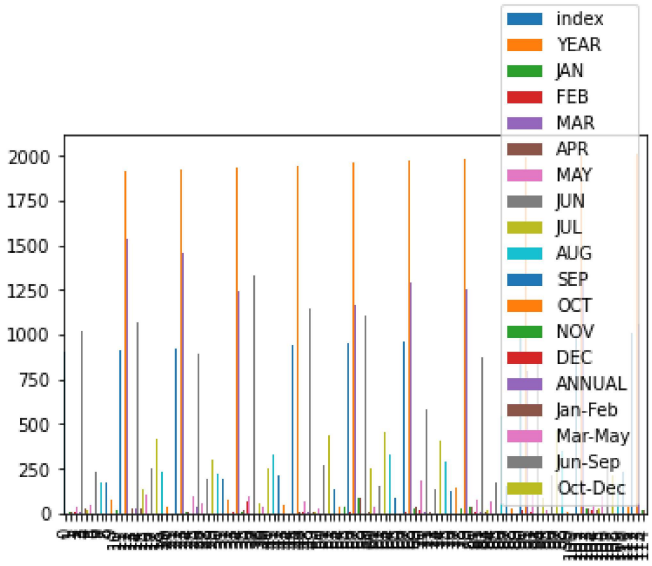
Out[7]: <AxesSubplot:>



Bar chart

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

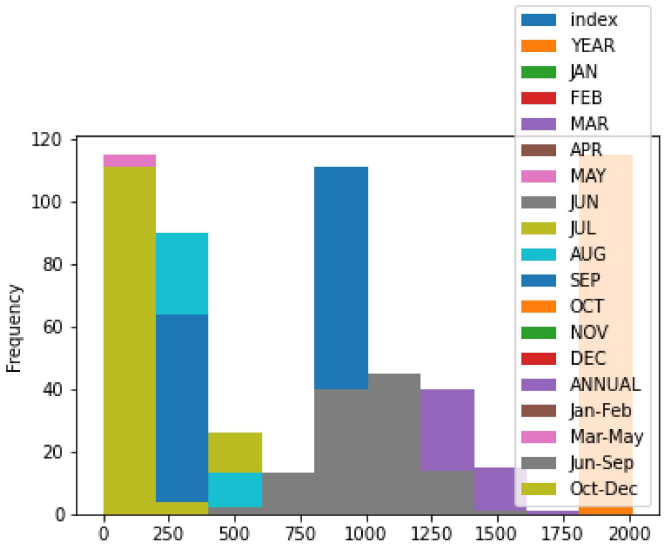
Out[8]: <AxesSubplot:>



Histogram

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

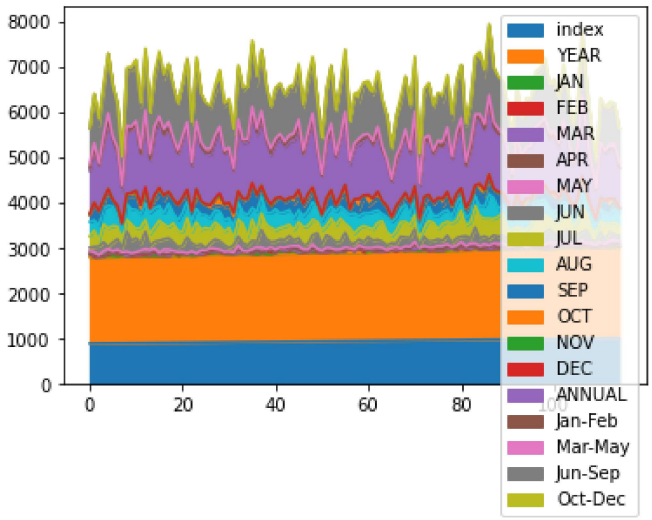
Out[9]:
```
<AxesSubplot:ylabel='Frequency'>
```



Area chart

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

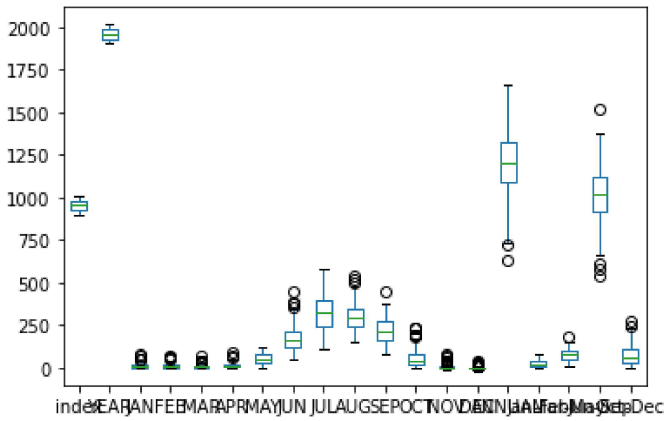
Out[10]:
```
<AxesSubplot:>
```



Box chart

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

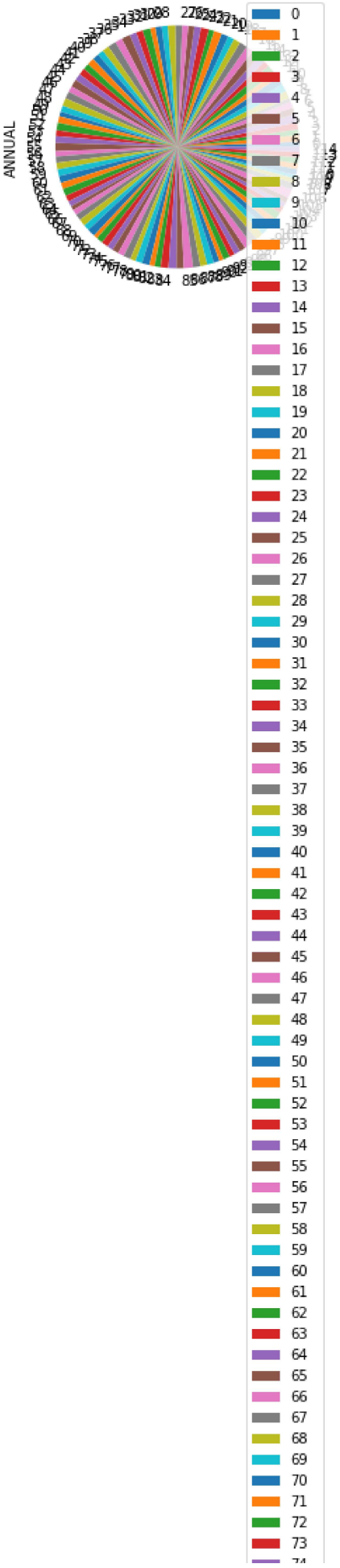
Out[11]: <AxesSubplot:>

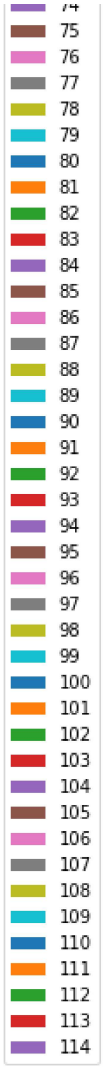


Pie chart

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

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

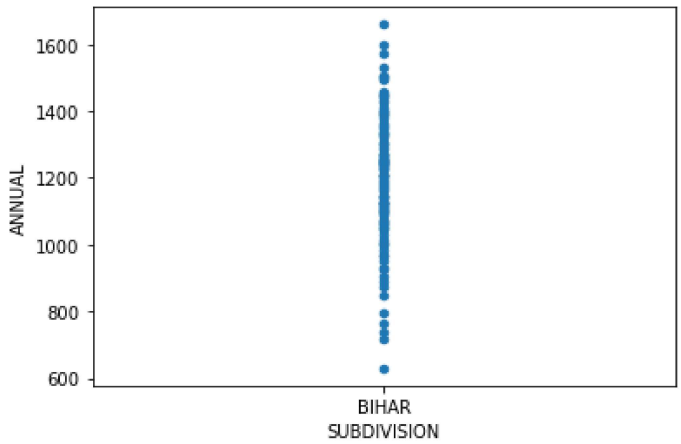

ANNUAL
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73



Scatter chart

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

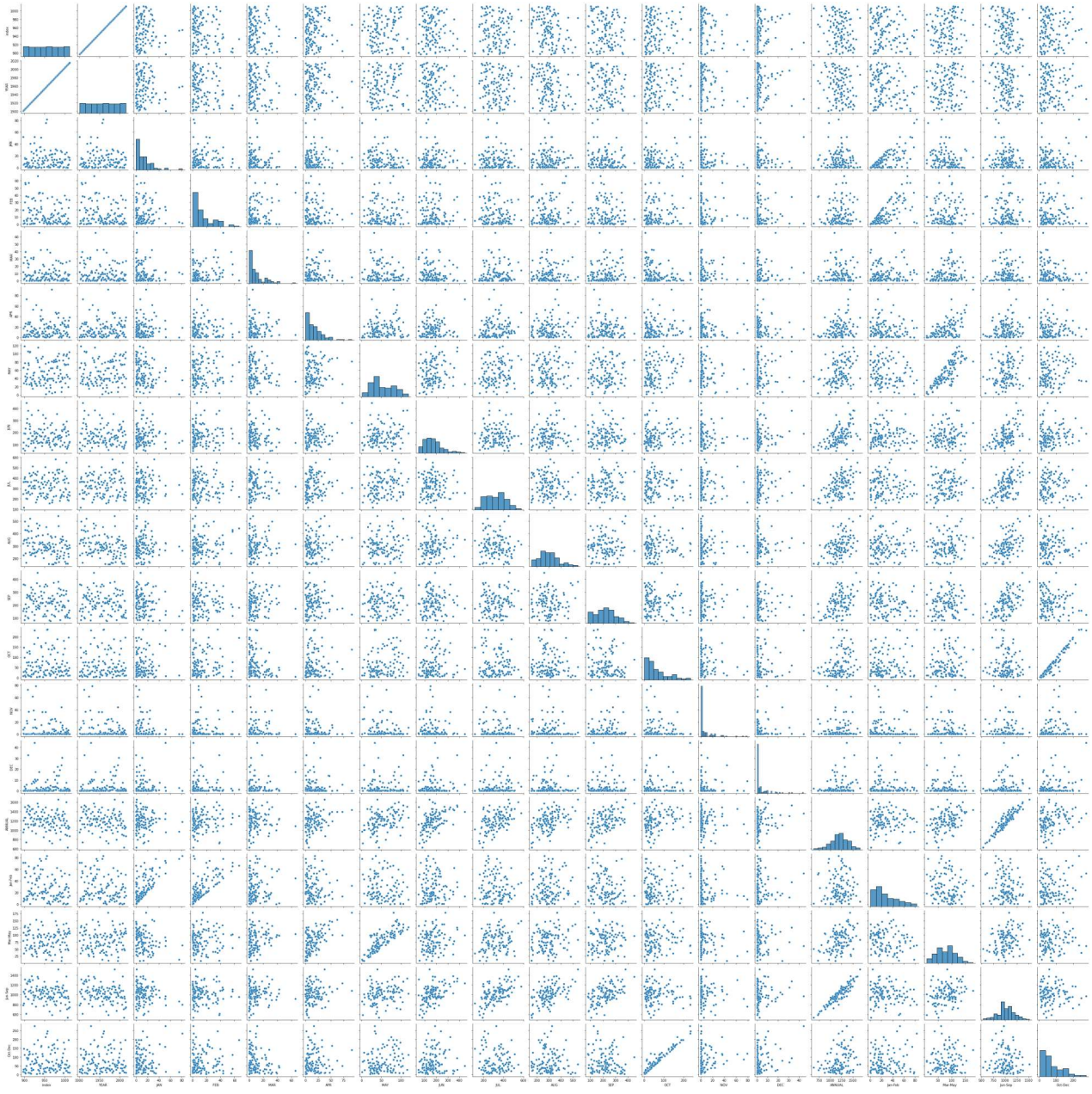
Out[13]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>



Seaborn

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

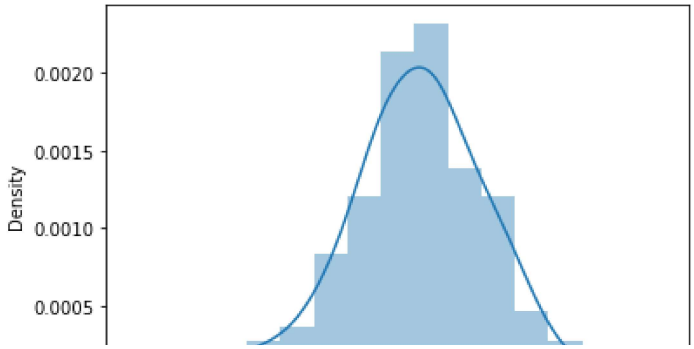
Out[14]:
```
<seaborn.axisgrid.PairGrid at 0x1f7c1639790>
```



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

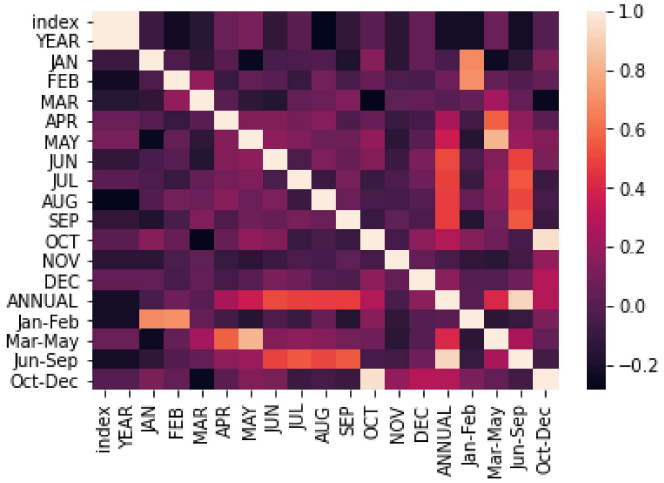
```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function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

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



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

Out[16]: <AxesSubplot:>



In [ ]:
```

```