tophw6dir

August 4, 2023

1 20104169 - SUMESH R

2 Importing Libraries

[1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

[2]:
```
from google.colab import drive
drive.mount('/content/drive')
df=pd.read_csv("/content/drive/MyDrive/mydatasets/rainfall/rainfall in india␣
↪1901-2015.csv")
df
```

```
Mounted at /content/drive
```

[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	\
0	0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	
1	1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	
2	2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	
3	3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	
4	4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	
…	…	…	…	…	…	…	…	…	
4111	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	
4112	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	
4113	4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	
4114	4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	
4115	4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	

	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	\
0	517.5	365.1	481.1	332.6	388.5	558.2	33.6	3373.2	136.3	
1	537.1	228.9	753.7	666.2	197.2	359.0	160.5	3520.7	159.8	
2	479.9	728.4	326.7	339.0	181.2	284.4	225.0	2957.4	156.7	
3	495.1	502.0	160.1	820.4	222.2	308.7	40.1	3079.6	24.1	
4	628.7	368.7	330.5	297.0	260.7	25.4	344.7	2566.7	1.3	
…	…	…	…	…	…	…	…	…	…	

```
4111  153.6  350.2  254.0  255.2  117.4  184.3   14.9  1533.7      7.9
4112  327.0  231.5  381.2  179.8  145.9   12.4    8.8  1405.5     19.3
4113  426.2  296.4  154.4  180.0   72.8   78.1   26.7  1426.3     60.6
4114  244.1  116.1  466.1  132.2  169.2   59.0   62.3  1395.0     69.3
4115  296.6  257.5  146.4  160.4  165.4  231.0  159.0  1642.9      2.7

      Mar-May  Jun-Sep  Oct-Dec
0       560.3   1696.3    980.3
1       458.3   2185.9    716.7
2       236.1   1874.0    690.6
3       506.9   1977.6    571.0
4       309.7   1624.9    630.8
…         …       …         …
4111    196.2   1013.0    316.6
4112     99.6   1119.5    167.1
4113    131.1   1057.0    177.6
4114     76.7    958.5    290.5
4115    223.9    860.9    555.4

[4116 rows x 20 columns]
```

3 Data Cleaning and Data Preprocessing

[4]:
```
df=df.dropna()
```

[5]:
```
df.columns
```

[5]:
```
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')
```

[6]:
```
df.info()
```

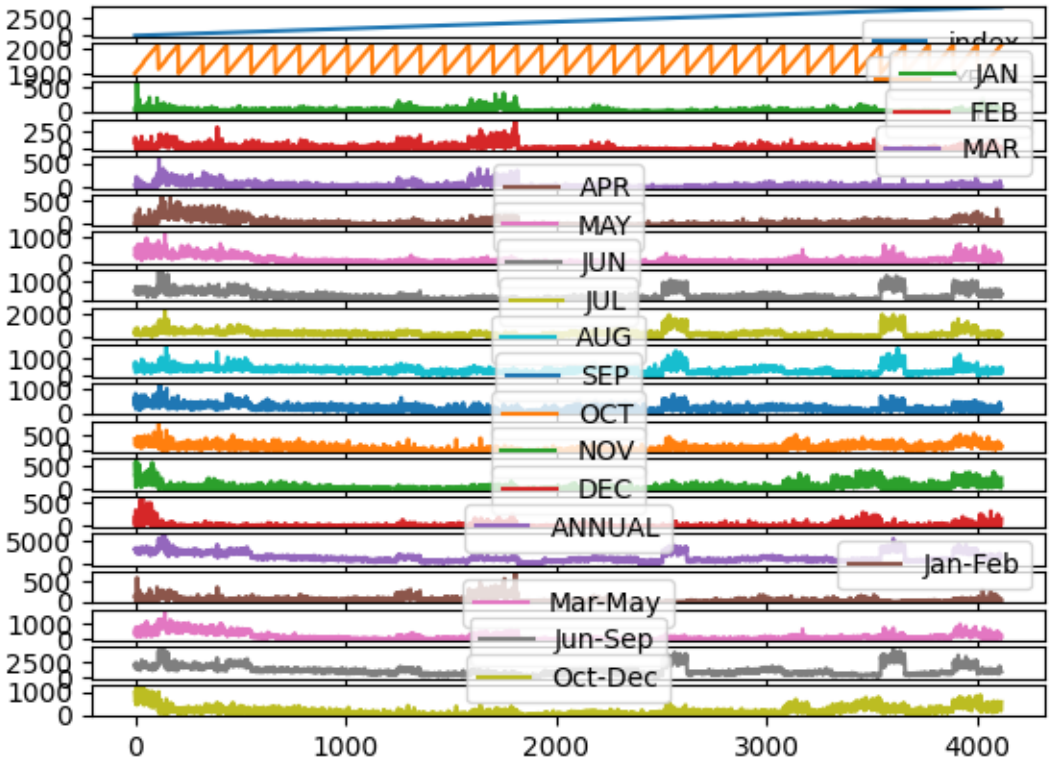
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4090 entries, 0 to 4115
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   index        4090 non-null   int64
 1   SUBDIVISION  4090 non-null   object
 2   YEAR         4090 non-null   int64
 3   JAN          4090 non-null   float64
 4   FEB          4090 non-null   float64
 5   MAR          4090 non-null   float64
 6   APR          4090 non-null   float64
```

```
 7   MAY          4090 non-null   float64
 8   JUN          4090 non-null   float64
 9   JUL          4090 non-null   float64
 10  AUG          4090 non-null   float64
 11  SEP          4090 non-null   float64
 12  OCT          4090 non-null   float64
 13  NOV          4090 non-null   float64
 14  DEC          4090 non-null   float64
 15  ANNUAL       4090 non-null   float64
 16  Jan-Feb      4090 non-null   float64
 17  Mar-May      4090 non-null   float64
 18  Jun-Sep      4090 non-null   float64
 19  Oct-Dec      4090 non-null   float64
dtypes: float64(17), int64(2), object(1)
memory usage: 671.0+ KB
```

4 Line chart

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

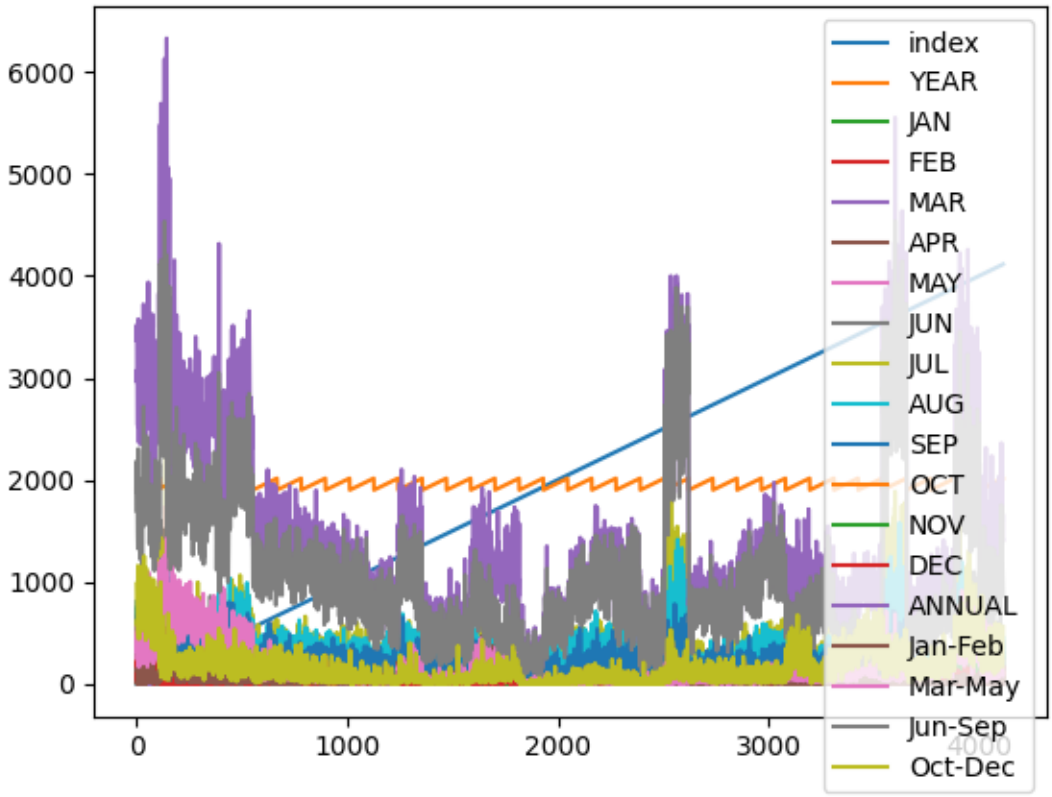
[9]:
```
array([<Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
       <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
       <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
       <Axes: >], dtype=object)
```



5 Line chart

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

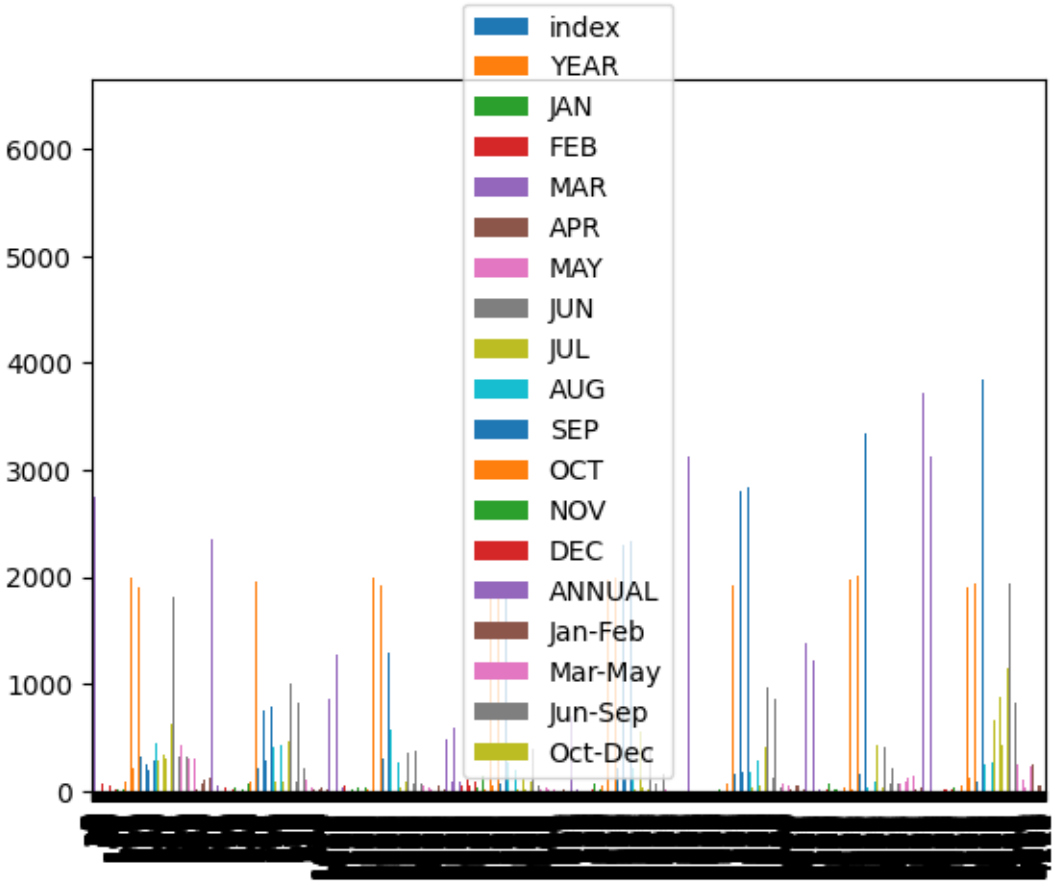
[10]: <Axes: >



6 Bar chart

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

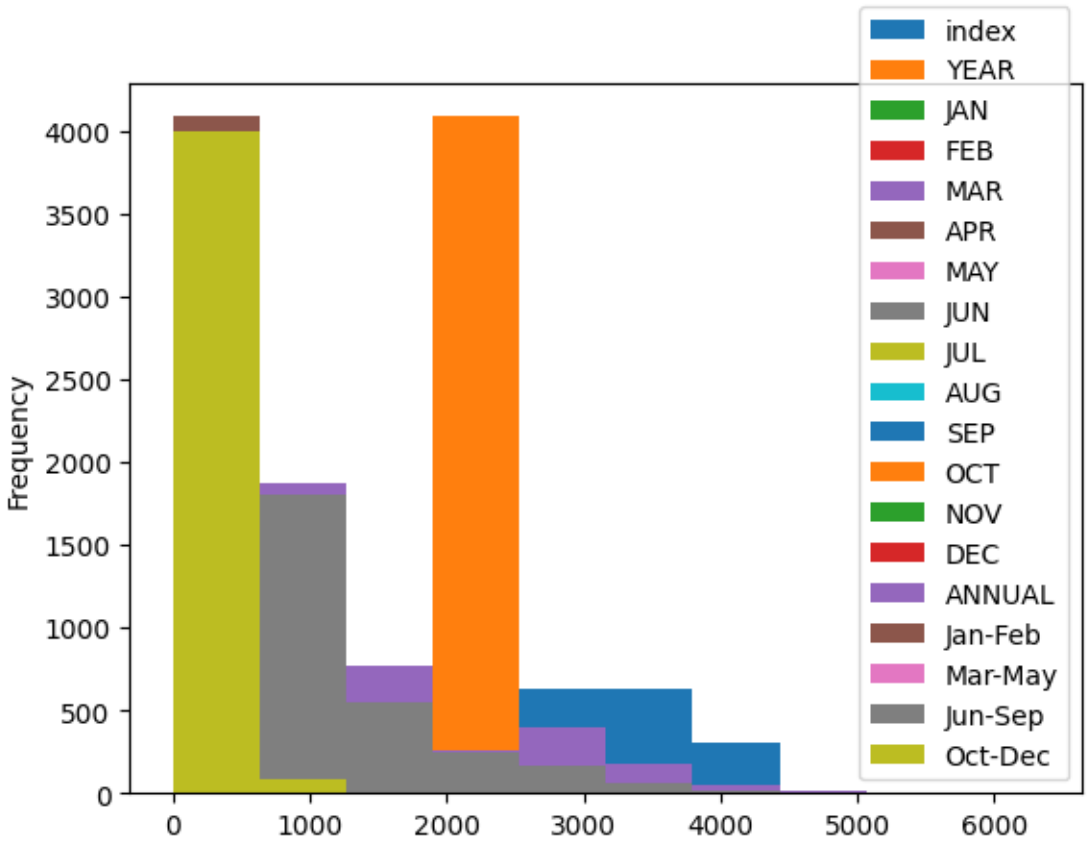
[11]: <Axes: >



7 Histogram

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

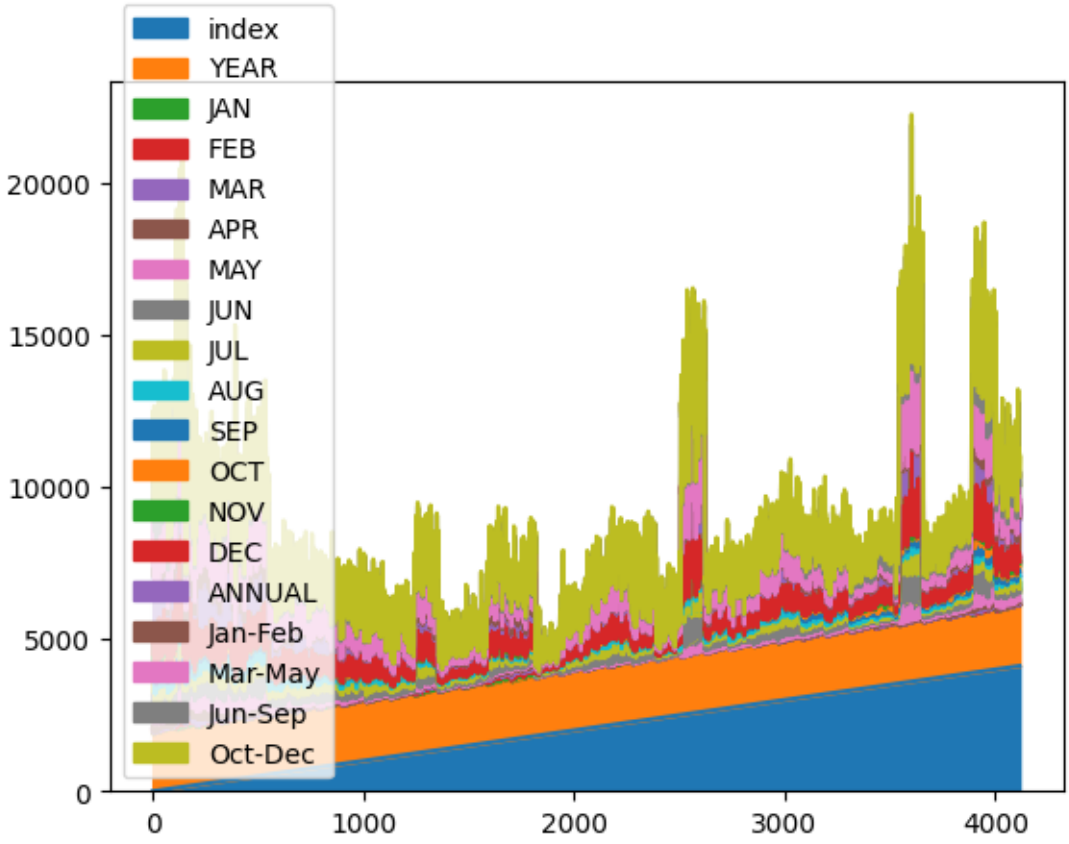
[12]: <Axes: ylabel='Frequency'>



8 Area chart

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

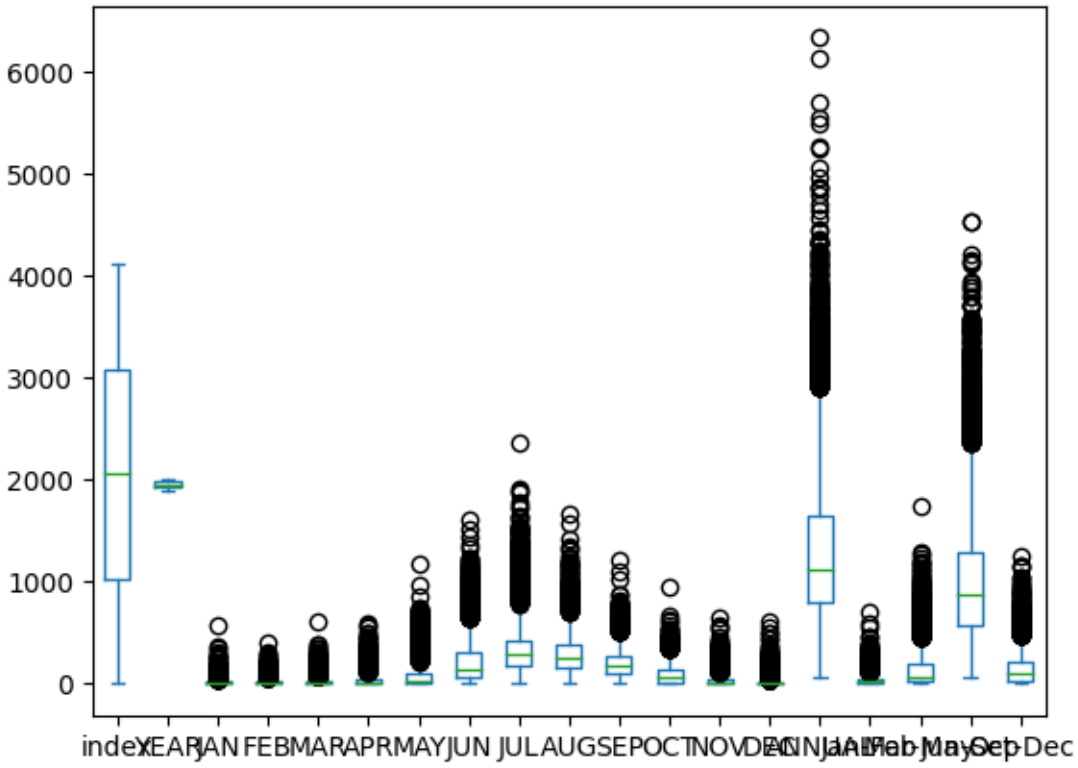
[13]: <Axes: >



9 Box chart

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

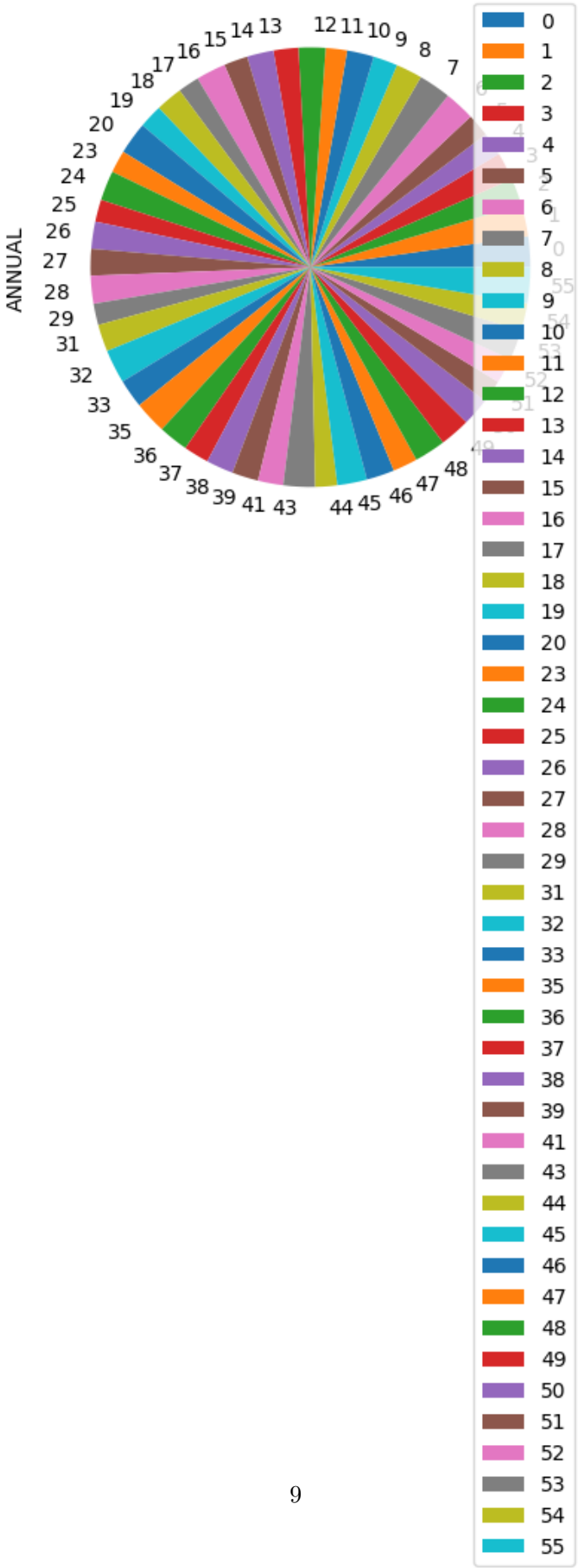
[14]: <Axes: >



10 Pie chart

[19]:
```
df[0:50].plot.pie(y='ANNUAL')
```

[19]:
```
<Axes: ylabel='ANNUAL'>
```


ANNUAL
12 11 10 9 8 7 6 5 4 3 2 1 0 55 54 53 52 51 49 48 47 46 45 44 43 41 39 38 37 36 35 33 32 31 29 28 27 26 25 24 23 20 19 18 17 16 15 14 13
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
23
24
25
26
27
28
29
31
32
33
35
36
37
38
39
41
43
44
45
46
47
48
49
50
51
52
53
54
55

11 Scatter chart

[15]:
```
df.plot.scatter(x='SUBDIVISION' ,y='ANNUAL')
```

[15]: <Axes: xlabel='SUBDIVISION', ylabel='ANNUAL'>