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

In [2]:
```
data=pd.read_csv(r"C:\Users\user\Desktop\vicky\rainfall\rainfall in india 1901-2015.csv")[4003:411
```

In [3]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113 entries, 4003 to 4115
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   index        113 non-null    int64  
 1   SUBDIVISION  113 non-null    object 
 2   YEAR         113 non-null    int64  
 3   JAN          111 non-null    float64
 4   FEB          112 non-null    float64
 5   MAR          111 non-null    float64
 6   APR          111 non-null    float64
 7   MAY          111 non-null    float64
 8   JUN          111 non-null    float64
 9   JUL          110 non-null    float64
 10  AUG          111 non-null    float64
 11  SEP          110 non-null    float64
 12  OCT          110 non-null    float64
 13  NOV          107 non-null    float64
 14  DEC          109 non-null    float64
 15  ANNUAL       102 non-null    float64
 16  Jan-Feb      110 non-null    float64
 17  Mar-May      109 non-null    float64
 18  Jun-Sep      109 non-null    float64
 19  Oct-Dec      107 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 17.8+ KB
```

In [4]:
```
data.head()
```

Out[4]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUA
4003	4003	LAKSHADWEEP	1902	99.3	9.6	32.6	40.4	179.1	374.2	413.3	170.0	214.3	384.2	192.8	49.0	2158
4004	4004	LAKSHADWEEP	1903	63.5	95.0	0.0	29.5	144.1	212.4	261.8	202.0	292.1	79.1	NaN	NaN	Na
4005	4005	LAKSHADWEEP	1904	0.0	0.0	13.5	13.2	143.3	261.3	256.0	38.9	219.9	153.6	8.3	68.9	1176
4006	4006	LAKSHADWEEP	1905	62.4	0.0	0.0	0.0	166.7	400.7	68.7	377.5	107.5	232.1	159.3	0.0	1574
4007	4007	LAKSHADWEEP	1906	17.8	0.0	24.4	33.8	213.0	465.0	348.6	260.5	25.9	252.3	106.5	63.8	1811

In [5]:
```
data.tail()
```

Out[5]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNU
4111	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2	117.4	184.3	14.9	153
4112	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8	145.9	12.4	8.8	140
4113	4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.0	72.8	78.1	26.7	142
4114	4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.2	169.2	59.0	62.3	139
4115	4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	296.6	257.5	146.4	160.4	165.4	231.0	159.0	164

In [6]:
```
data.shape
```

Out[6]:
```
(113, 20)
```

In [7]:
```
new_data=data.fillna(value=1)
new_data
```

Out[7]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNU
4003	4003	LAKSHADWEEP	1902	99.3	9.6	32.6	40.4	179.1	374.2	413.3	170.0	214.3	384.2	192.8	49.0	215
4004	4004	LAKSHADWEEP	1903	63.5	95.0	0.0	29.5	144.1	212.4	261.8	202.0	292.1	79.1	1.0	1.0	
4005	4005	LAKSHADWEEP	1904	0.0	0.0	13.5	13.2	143.3	261.3	256.0	38.9	219.9	153.6	8.3	68.9	117
4006	4006	LAKSHADWEEP	1905	62.4	0.0	0.0	0.0	166.7	400.7	68.7	377.5	107.5	232.1	159.3	0.0	157
4007	4007	LAKSHADWEEP	1906	17.8	0.0	24.4	33.8	213.0	465.0	348.6	260.5	25.9	252.3	106.5	63.8	181
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
4111	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2	117.4	184.3	14.9	153
4112	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8	145.9	12.4	8.8	140
4113	4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.0	72.8	78.1	26.7	142
4114	4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.2	169.2	59.0	62.3	139
4115	4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	296.6	257.5	146.4	160.4	165.4	231.0	159.0	164

113 rows × 20 columns

In [8]:
```
new_data.index
```

Out[8]:
```
RangeIndex(start=4003, stop=4116, step=1)
```

In [9]:
```
new_data.columns
```

Out[9]:
```
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 [10]:
```
new_data.plot.line()
```

Out[10]: <AxesSubplot:>

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

Out[11]: <AxesSubplot:>

In [12]:
```
new_data.plot.area()
```

Out[12]:
```
<AxesSubplot:>
```

In [13]:
```
new_data.plot.hist()
```

Out[13]:
```
<AxesSubplot:ylabel='Frequency'>
```

In [14]:
```
new_data.plot.pie(y='ANNUAL')
```

Out[14]:
```
<AxesSubplot:ylabel='ANNUAL'>
```

In [15]:
```
new_data.plot.scatter(x='YEAR',y='ANNUAL')
```

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

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

Out[16]:
```
<seaborn.axisgrid.PairGrid at 0x25acc843820>
```

In [17]:
```
sns.distplot(data['YEAR'])
```

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

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

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

Out[18]: <AxesSubplot:>