Import Libraies

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

In [2]:
```
df=pd.read_csv(r"C:\Users\user\Downloads\FP2_RainFall\rainfall in india 1901-2015.csv")[1127:124
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
1127	1127	WEST UTTAR PRADESH	1901	51.4	25.6	9.5	0.7	5.6	23.8	201.9	374.3	67.7	7.6	0.0	7.9	775.9
1128	1128	WEST UTTAR PRADESH	1902	4.6	4.6	0.6	4.8	7.2	54.5	325.9	180.6	143.1	9.6	0.9	0.2	736.6
1129	1129	WEST UTTAR PRADESH	1903	13.4	0.4	1.2	0.0	8.2	32.7	145.4	279.1	150.4	177.3	0.0	0.4	808.4
1130	1130	WEST UTTAR PRADESH	1904	6.3	2.0	29.7	0.4	24.8	68.5	358.8	311.1	97.1	2.7	15.7	28.2	945.2
1131	1131	WEST UTTAR PRADESH	1905	32.3	26.6	14.8	3.6	7.1	18.9	139.8	95.0	92.2	0.2	0.0	2.9	433.5
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1236	1236	WEST UTTAR PRADESH	2010	3.2	13.7	0.2	0.4	9.3	12.5	282.0	213.8	263.9	4.7	12.8	4.1	820.7
1237	1237	WEST UTTAR PRADESH	2011	2.1	10.4	3.9	2.8	29.6	175.9	215.9	232.3	101.7	0.7	0.5	1.5	777.4
1238	1238	WEST UTTAR PRADESH	2012	14.5	0.1	1.4	4.7	0.3	4.0	145.1	149.1	67.8	0.5	0.1	2.0	389.6
1239	1239	WEST UTTAR PRADESH	2013	20.4	69.5	3.5	1.6	2.1	190.6	233.9	287.1	52.2	61.2	1.7	8.9	932.8
1240	1240	WEST UTTAR PRADESH	2014	48.3	29.4	22.6	5.3	11.0	22.0	151.6	81.0	84.7	14.6	0.0	16.3	486.9

114 rows × 20 columns

Data Cleaning and Preprocessing

In [3]:
```
df.dropna()
```

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
1127	1127	WEST UTTAR PRADESH	1901	51.4	25.6	9.5	0.7	5.6	23.8	201.9	374.3	67.7	7.6	0.0	7.9	775.9
1128	1128	WEST UTTAR PRADESH	1902	4.6	4.6	0.6	4.8	7.2	54.5	325.9	180.6	143.1	9.6	0.9	0.2	736.6
1129	1129	WEST UTTAR PRADESH	1903	13.4	0.4	1.2	0.0	8.2	32.7	145.4	279.1	150.4	177.3	0.0	0.4	808.4
1130	1130	WEST UTTAR PRADESH	1904	6.3	2.0	29.7	0.4	24.8	68.5	358.8	311.1	97.1	2.7	15.7	28.2	945.2
1131	1131	WEST UTTAR PRADESH	1905	32.3	26.6	14.8	3.6	7.1	18.9	139.8	95.0	92.2	0.2	0.0	2.9	433.5
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1236	1236	WEST UTTAR PRADESH	2010	3.2	13.7	0.2	0.4	9.3	12.5	282.0	213.8	263.9	4.7	12.8	4.1	820.7
1237	1237	WEST UTTAR PRADESH	2011	2.1	10.4	3.9	2.8	29.6	175.9	215.9	232.3	101.7	0.7	0.5	1.5	777.4
1238	1238	WEST UTTAR PRADESH	2012	14.5	0.1	1.4	4.7	0.3	4.0	145.1	149.1	67.8	0.5	0.1	2.0	389.6
1239	1239	WEST UTTAR PRADESH	2013	20.4	69.5	3.5	1.6	2.1	190.6	233.9	287.1	52.2	61.2	1.7	8.9	932.8
1240	1240	WEST UTTAR PRADESH	2014	48.3	29.4	22.6	5.3	11.0	22.0	151.6	81.0	84.7	14.6	0.0	16.3	486.9

114 rows × 20 columns

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')
```

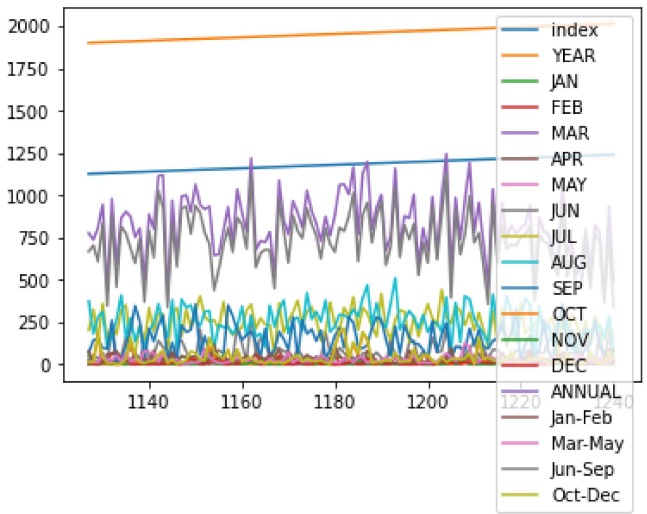
In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 114 entries, 1127 to 1240
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   index        114 non-null    int64  
 1   SUBDIVISION  114 non-null    object 
 2   YEAR         114 non-null    int64  
 3   JAN          114 non-null    float64
 4   FEB          114 non-null    float64
 5   MAR          114 non-null    float64
 6   APR          114 non-null    float64
 7   MAY          114 non-null    float64
 8   JUN          114 non-null    float64
 9   JUL          114 non-null    float64
 10  AUG          114 non-null    float64
 11  SEP          114 non-null    float64
 12  OCT          114 non-null    float64
 13  NOV          114 non-null    float64
 14  DEC          114 non-null    float64
 15  ANNUAL       114 non-null    float64
 16  Jan-Feb      114 non-null    float64
 17  Mar-May      114 non-null    float64
 18  Jun-Sep      114 non-null    float64
 19  Oct-Dec      114 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 17.9+ KB
```

Line chart

In [6]:
```
df.plot.line()
```

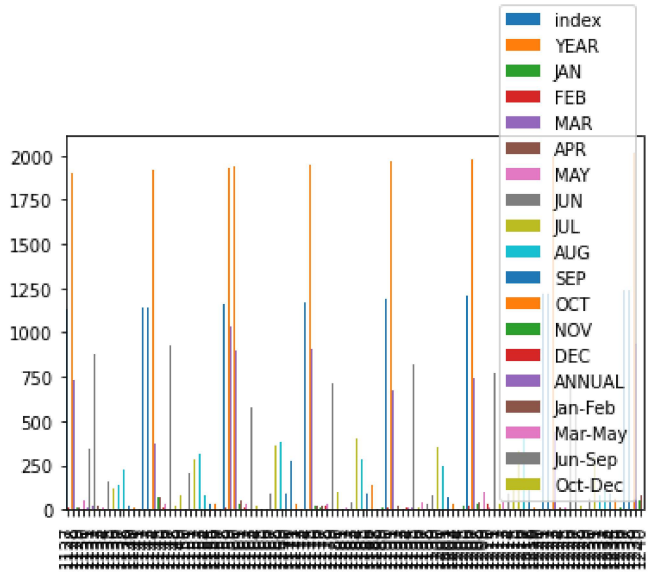
Out[6]: <AxesSubplot:>



Bar chart

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

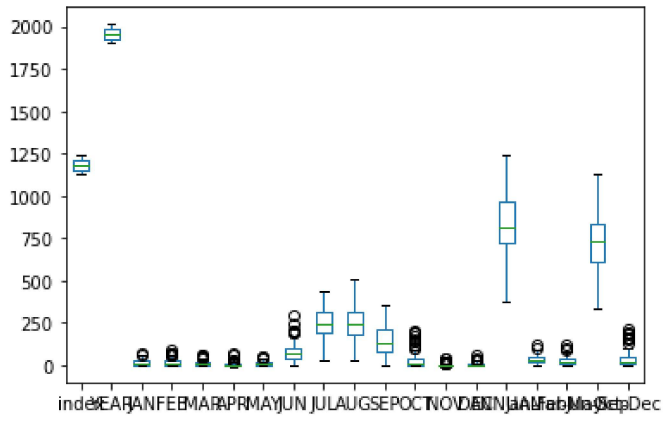
Out[7]: <AxesSubplot:>



Box chart

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

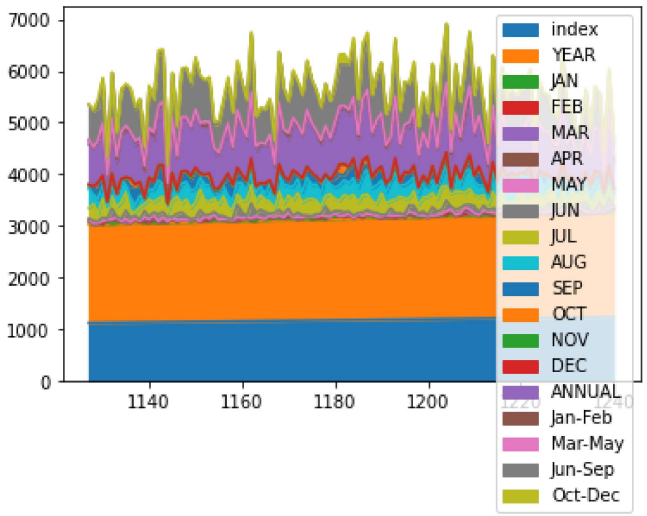
Out[8]: <AxesSubplot:>



Area chart

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

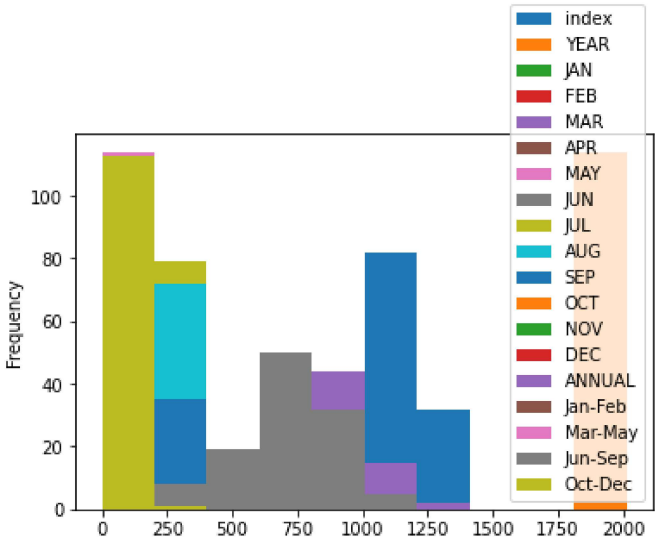
Out[9]: <AxesSubplot:>



Histogram

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

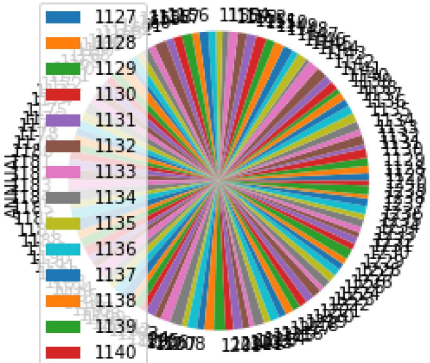
Out[10]: <AxesSubplot:ylabel='Frequency'>

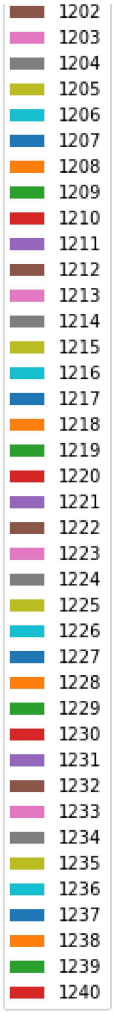


pie chart

In [11]:
```
df.plot.pie(y="ANNUAL")
```

Out[11]:
```
<AxesSubplot:ylabel='ANNUAL'>
```

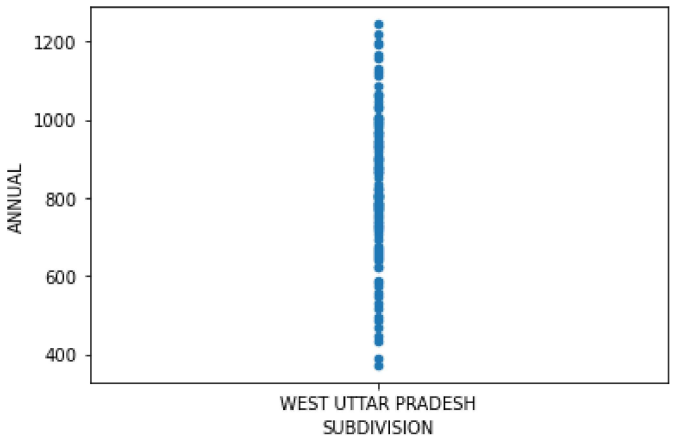

1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201



Scatter chart

In [12]:
```
df.plot.scatter(y='ANNUAL',x='SUBDIVISION')
```

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



In [13]:
```
df.describe()
```

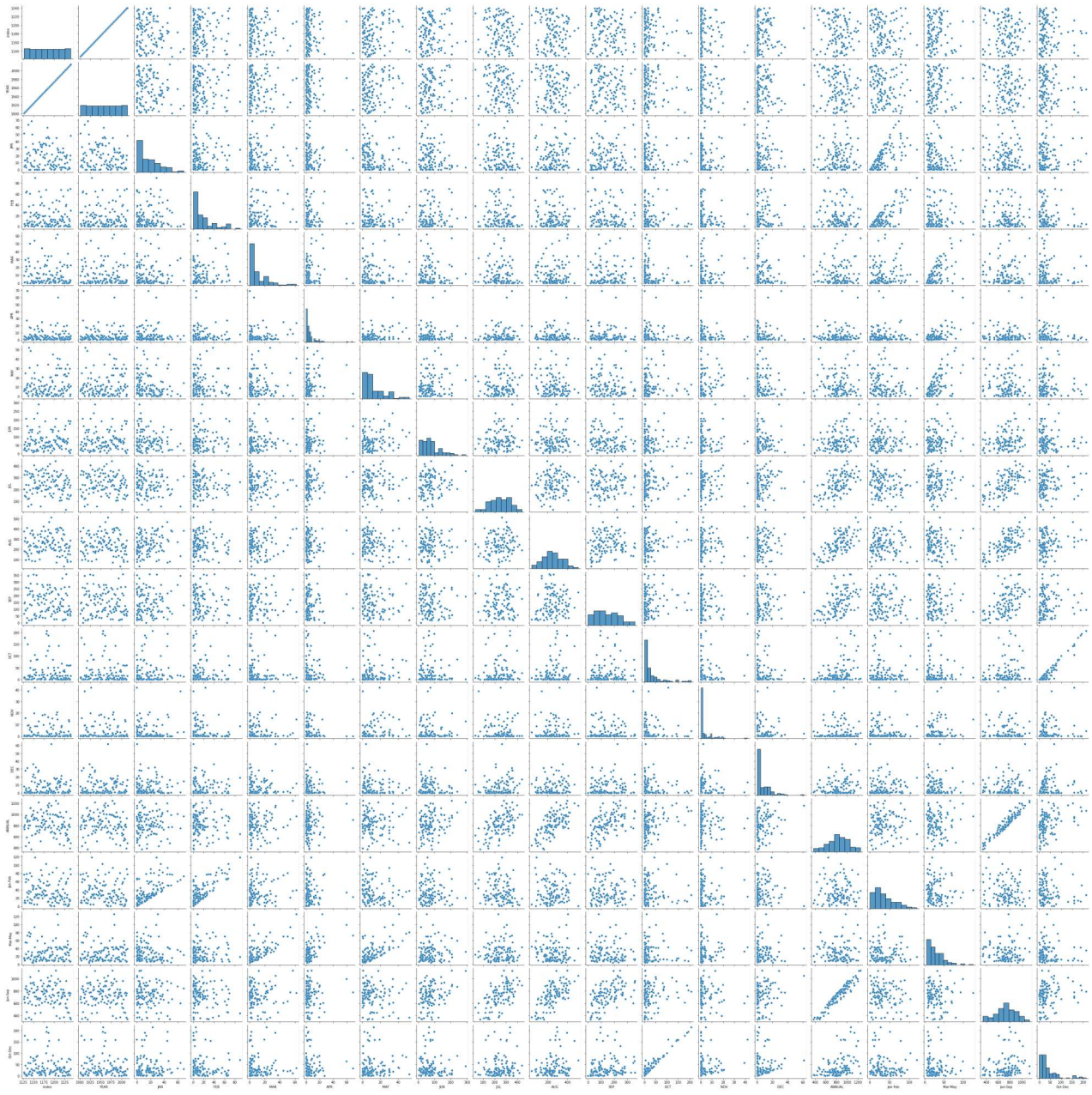
Out[13]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	
count	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	1
mean	1183.500000	1957.500000	17.543860	17.987719	10.976316	6.123684	12.342982	77.646491	246.978947	2
std	33.052988	33.052988	15.806518	20.035505	13.363027	9.962810	11.572587	55.925154	84.657438	
min	1127.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	3.700000	29.600000	
25%	1155.250000	1929.250000	4.050000	3.225000	1.700000	0.725000	3.725000	39.200000	197.375000	1
50%	1183.500000	1957.500000	14.000000	10.800000	5.600000	3.000000	8.150000	67.150000	241.200000	2
75%	1211.750000	1985.750000	27.650000	25.600000	17.050000	6.425000	18.700000	97.325000	311.400000	3
max	1240.000000	2014.000000	68.600000	89.900000	61.800000	69.000000	52.000000	291.200000	440.600000	5

EDA AND VISUALIZATION

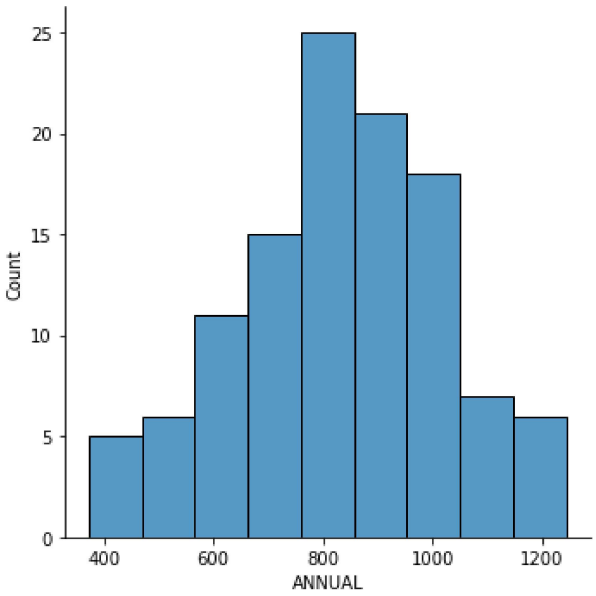
In [14]:
```
sns.pairplot(df)
```

Out[14]:
```
<seaborn.axisgrid.PairGrid at 0x20671c57490>
```



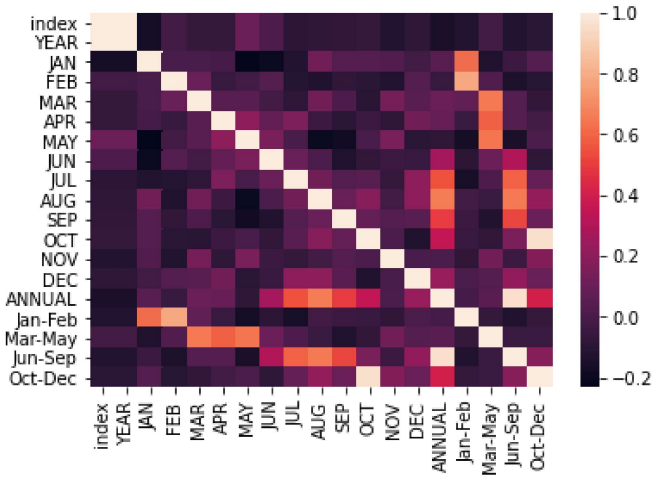
In [15]:
```
sns.displot(df['ANNUAL'])
```

Out[15]: <seaborn.axisgrid.FacetGrid at 0x206002cdca0>



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

Out[16]: <AxesSubplot:>



In [ ]:
```

```