

# 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")[1242:1356]
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNI
1242	1242	UTTARAKHAND	1901	134.5	81.4	44.5	5.9	60.8	33.6	381.1	612.3	167.1	16.3	0.0	24.9	15
1243	1243	UTTARAKHAND	1902	0.0	17.0	52.2	63.7	52.1	113.1	444.1	327.5	220.4	31.9	2.1	0.0	13
1244	1244	UTTARAKHAND	1903	68.0	7.9	87.6	10.3	37.5	83.0	251.6	442.7	249.3	57.5	0.0	11.3	13
1245	1245	UTTARAKHAND	1904	40.0	5.2	78.3	13.6	61.1	180.1	449.6	417.2	174.1	6.3	35.6	31.0	14
1246	1246	UTTARAKHAND	1905	115.4	80.7	99.8	26.1	70.3	111.5	299.9	349.5	129.5	0.0	1.0	18.5	13
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1351	1351	UTTARAKHAND	2010	11.4	62.1	3.8	13.8	47.0	72.7	455.8	335.7	389.7	8.0	4.5	19.5	14
1352	1352	UTTARAKHAND	2011	30.9	65.2	18.0	30.9	84.2	223.1	433.3	523.7	148.4	3.4	1.2	2.3	15
1353	1353	UTTARAKHAND	2012	38.8	11.9	28.1	39.2	9.1	46.0	387.1	419.5	220.6	4.7	3.4	15.5	12
1354	1354	UTTARAKHAND	2013	73.0	188.3	22.0	24.7	18.2	488.9	413.4	359.4	111.3	29.1	3.2	3.8	17
1355	1355	UTTARAKHAND	2014	45.9	99.9	68.4	37.6	52.9	62.9	462.7	264.2	107.9	40.8	0.0	44.3	12

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	ANNI
1242	1242	UTTARAKHAND	1901	134.5	81.4	44.5	5.9	60.8	33.6	381.1	612.3	167.1	16.3	0.0	24.9	15
1243	1243	UTTARAKHAND	1902	0.0	17.0	52.2	63.7	52.1	113.1	444.1	327.5	220.4	31.9	2.1	0.0	13
1244	1244	UTTARAKHAND	1903	68.0	7.9	87.6	10.3	37.5	83.0	251.6	442.7	249.3	57.5	0.0	11.3	13
1245	1245	UTTARAKHAND	1904	40.0	5.2	78.3	13.6	61.1	180.1	449.6	417.2	174.1	6.3	35.6	31.0	14
1246	1246	UTTARAKHAND	1905	115.4	80.7	99.8	26.1	70.3	111.5	299.9	349.5	129.5	0.0	1.0	18.5	13
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1351	1351	UTTARAKHAND	2010	11.4	62.1	3.8	13.8	47.0	72.7	455.8	335.7	389.7	8.0	4.5	19.5	14
1352	1352	UTTARAKHAND	2011	30.9	65.2	18.0	30.9	84.2	223.1	433.3	523.7	148.4	3.4	1.2	2.3	15
1353	1353	UTTARAKHAND	2012	38.8	11.9	28.1	39.2	9.1	46.0	387.1	419.5	220.6	4.7	3.4	15.5	12
1354	1354	UTTARAKHAND	2013	73.0	188.3	22.0	24.7	18.2	488.9	413.4	359.4	111.3	29.1	3.2	3.8	17
1355	1355	UTTARAKHAND	2014	45.9	99.9	68.4	37.6	52.9	62.9	462.7	264.2	107.9	40.8	0.0	44.3	12

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')

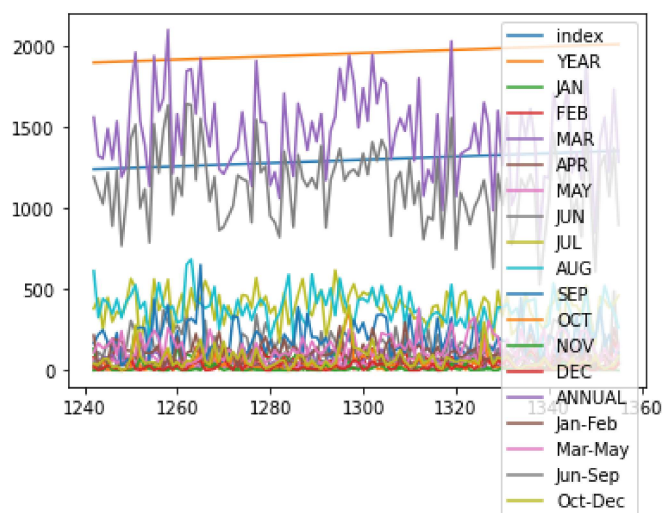
In [5]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 114 entries, 1242 to 1355
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()
```

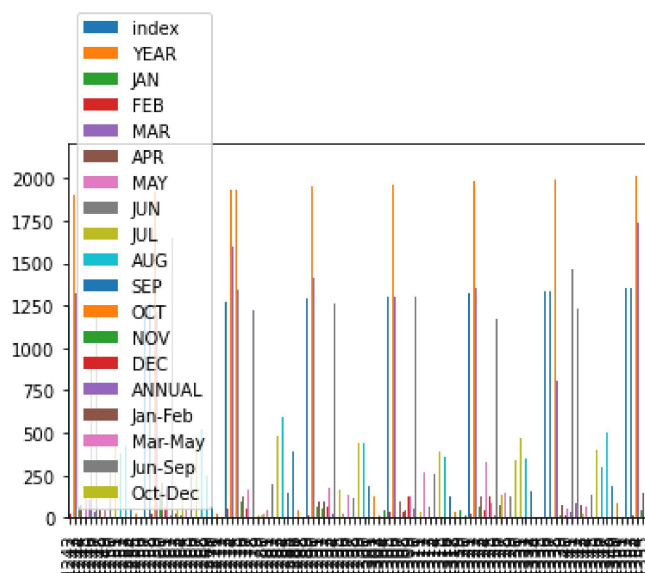
```
Out[6]: <AxesSubplot:>
```



## Bar chart

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

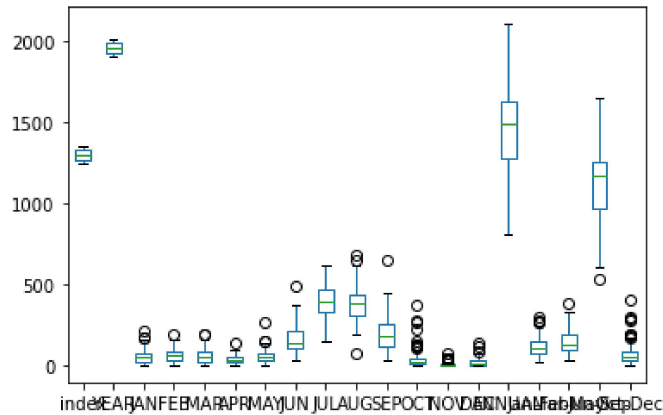
```
Out[7]: <AxesSubplot:>
```



## Box chart

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

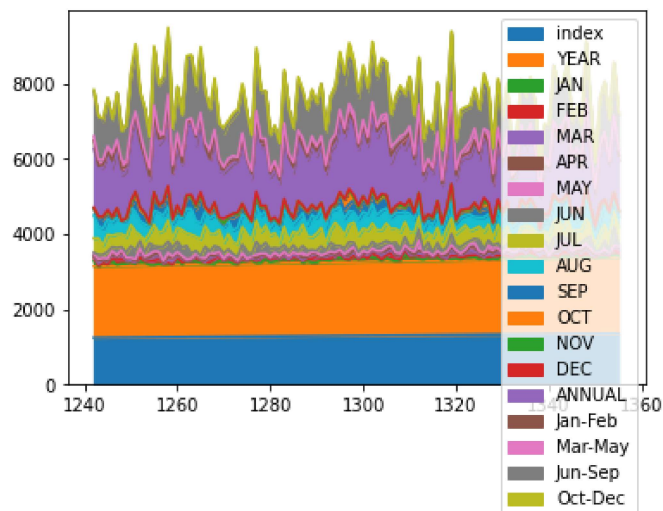
```
Out[8]: <AxesSubplot:>
```



## Area chart

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

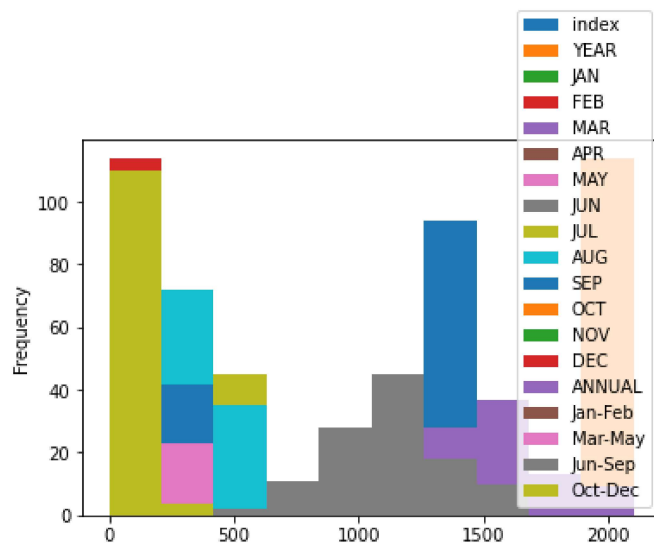
```
Out[9]: <AxesSubplot:>
```



## Histogram

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

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

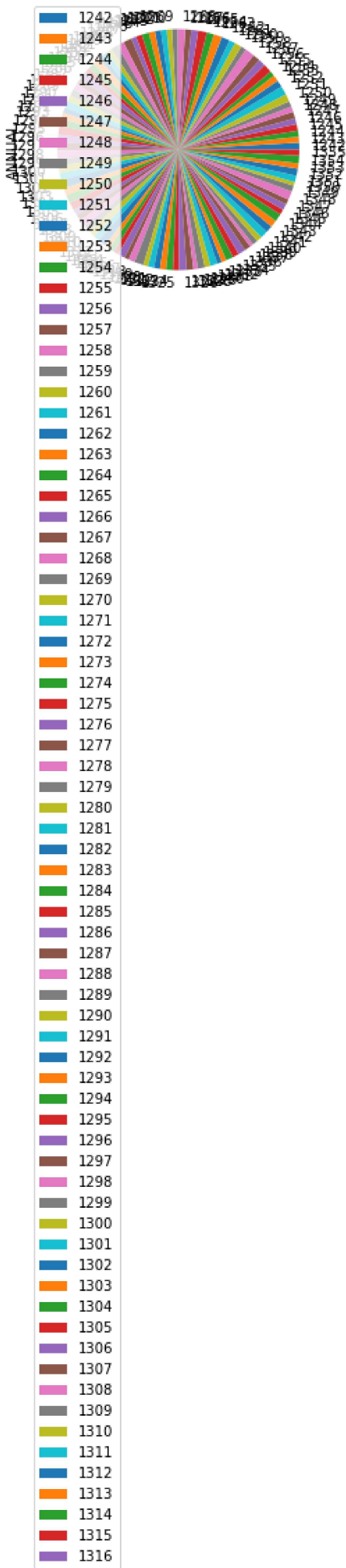


**pie chart**

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

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





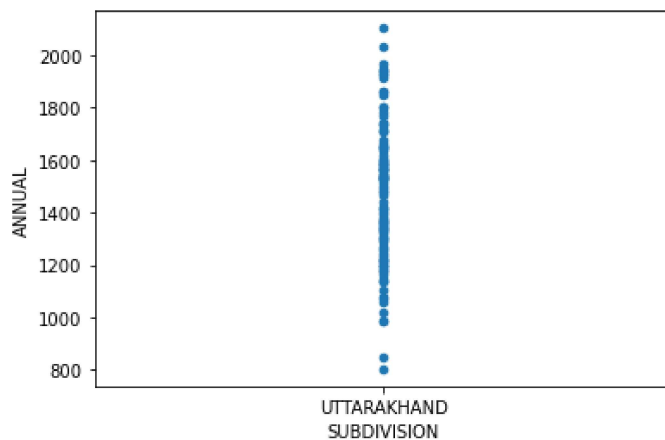


1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355

## Scatter chart

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

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



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

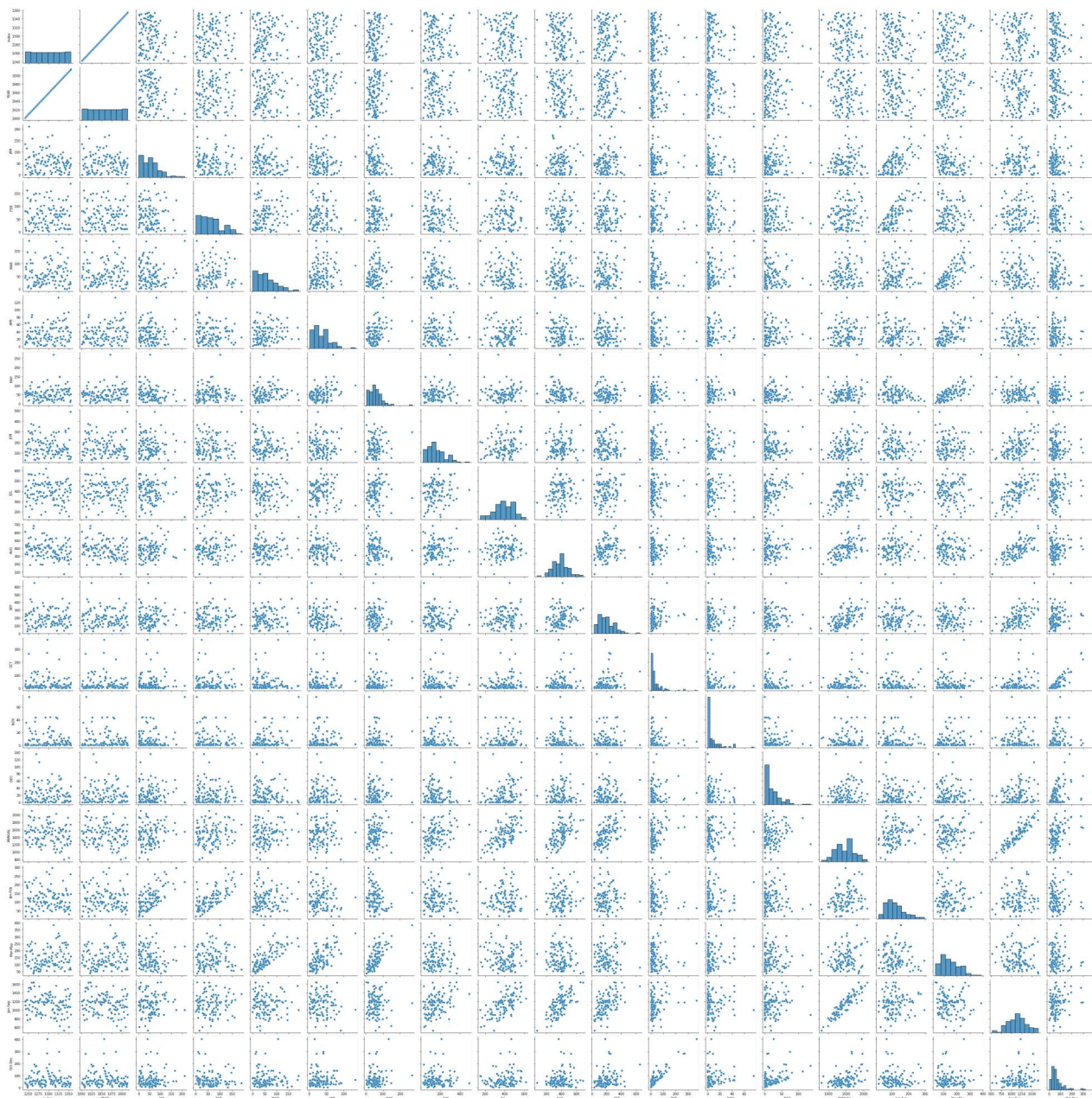
Out[13]:

	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
mean	1298.500000	1957.500000	53.791228	63.459649	56.657895	34.971930	55.490351	162.340351	391.169298
std	33.052988	33.052988	41.067849	44.234899	42.109478	24.132571	36.722979	86.622614	98.663373
min	1242.000000	1901.000000	0.000000	0.000000	0.000000	1.100000	3.600000	33.600000	149.300000
25%	1270.250000	1929.250000	21.250000	27.925000	22.575000	18.175000	27.825000	104.250000	330.175000
50%	1298.500000	1957.500000	49.550000	59.350000	47.250000	29.850000	50.750000	139.750000	394.350000
75%	1326.750000	1985.750000	76.400000	88.850000	79.500000	51.025000	71.575000	212.375000	466.625000
max	1355.000000	2014.000000	211.400000	188.300000	190.300000	132.900000	270.200000	488.900000	617.500000

# EDA AND VISUALIZATION

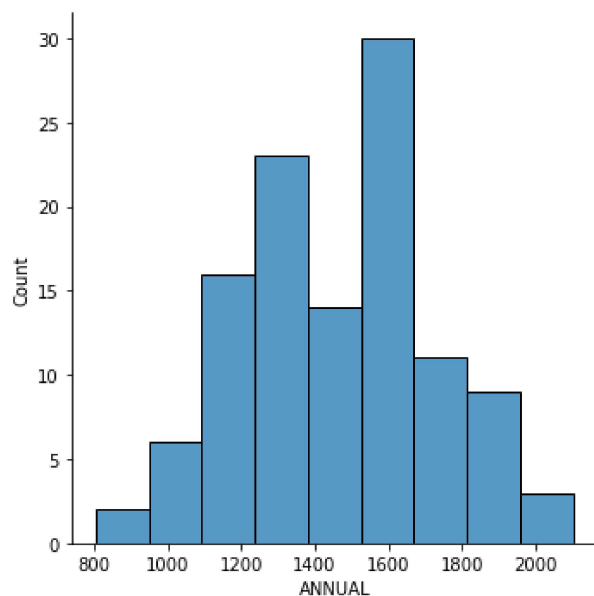
```
In [14]: sns.pairplot(df)
```

```
Out[14]: <seaborn.axisgrid.PairGrid at 0x142d6365760>
```



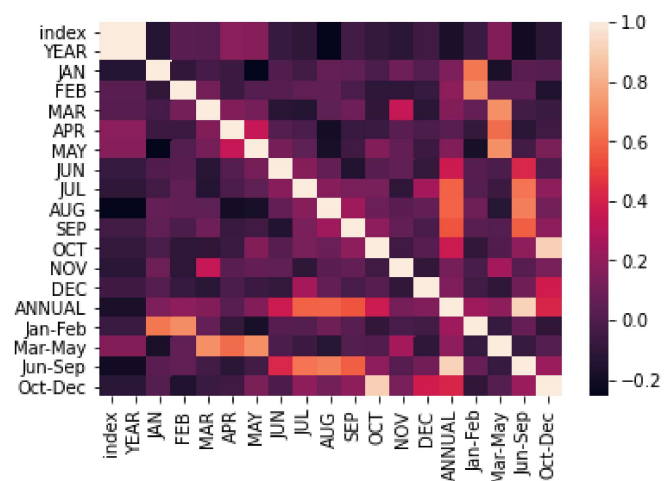
```
In [15]: sns.displot(df['ANNUAL'])
```

```
Out[15]: <seaborn.axisgrid.FacetGrid at 0x142e46da040>
```



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

```
Out[16]: <AxesSubplot:>
```



```
In [ ]:
```