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")[3427:354
        df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
3427	3427	TAMIL NADU	1901	24.5	39.1	21.7	36.0	74.0	41.8	49.3	67.9	191.1	122.3	212.3	80.4	960.3
3428	3428	TAMIL NADU	1902	67.2	9.8	25.1	21.9	84.7	39.3	55.1	113.8	98.6	282.2	174.9	165.8	1138.2
3429	3429	TAMIL NADU	1903	19.3	7.8	1.7	18.2	128.5	58.5	72.6	115.0	210.4	128.1	200.5	203.2	1163.9
3430	3430	TAMIL NADU	1904	35.2	0.1	0.7	19.5	121.9	34.9	89.0	40.4	85.7	163.2	23.6	49.1	663.1
3431	3431	TAMIL NADU	1905	6.5	7.5	17.2	64.8	83.7	49.8	39.0	101.8	73.5	250.4	123.7	3.2	821.1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..
3536	3536	TAMIL NADU	2010	11.8	0.2	1.9	22.9	91.9	70.0	81.4	102.9	111.1	148.1	328.6	124.5	1095.2
3537	3537	TAMIL NADU	2011	4.3	11.2	8.0	91.5	33.4	56.0	45.5	128.9	76.0	200.4	230.5	41.0	926.5
3538	3538	TAMIL NADU	2012	3.0	0.1	2.5	35.5	41.9	30.1	46.5	98.0	84.9	235.2	44.5	14.0	636.1
3539	3539	TAMIL NADU	2013	3.9	30.9	30.0	20.3	42.0	54.6	42.7	110.7	113.5	127.9	112.3	53.2	741.9
3540	3540	TAMIL NADU	2014	7.4	6.1	8.1	8.3	139.1	47.8	50.6	117.7	98.9	252.2	110.8	66.0	913.0

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
3427	3427	TAMIL NADU	1901	24.5	39.1	21.7	36.0	74.0	41.8	49.3	67.9	191.1	122.3	212.3	80.4	960.3
3428	3428	TAMIL NADU	1902	67.2	9.8	25.1	21.9	84.7	39.3	55.1	113.8	98.6	282.2	174.9	165.8	1138.2
3429	3429	TAMIL NADU	1903	19.3	7.8	1.7	18.2	128.5	58.5	72.6	115.0	210.4	128.1	200.5	203.2	1163.9
3430	3430	TAMIL NADU	1904	35.2	0.1	0.7	19.5	121.9	34.9	89.0	40.4	85.7	163.2	23.6	49.1	663.7
3431	3431	TAMIL NADU	1905	6.5	7.5	17.2	64.8	83.7	49.8	39.0	101.8	73.5	250.4	123.7	3.2	821.7
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3536	3536	TAMIL NADU	2010	11.8	0.2	1.9	22.9	91.9	70.0	81.4	102.9	111.1	148.1	328.6	124.5	1095.2
3537	3537	TAMIL NADU	2011	4.3	11.2	8.0	91.5	33.4	56.0	45.5	128.9	76.0	200.4	230.5	41.0	926.5
3538	3538	TAMIL NADU	2012	3.0	0.1	2.5	35.5	41.9	30.1	46.5	98.0	84.9	235.2	44.5	14.0	636.7
3539	3539	TAMIL NADU	2013	3.9	30.9	30.0	20.3	42.0	54.6	42.7	110.7	113.5	127.9	112.3	53.2	741.9
3540	3540	TAMIL NADU	2014	7.4	6.1	8.1	8.3	139.1	47.8	50.6	117.7	98.9	252.2	110.8	66.0	913.0

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, 3427 to 3540
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
```
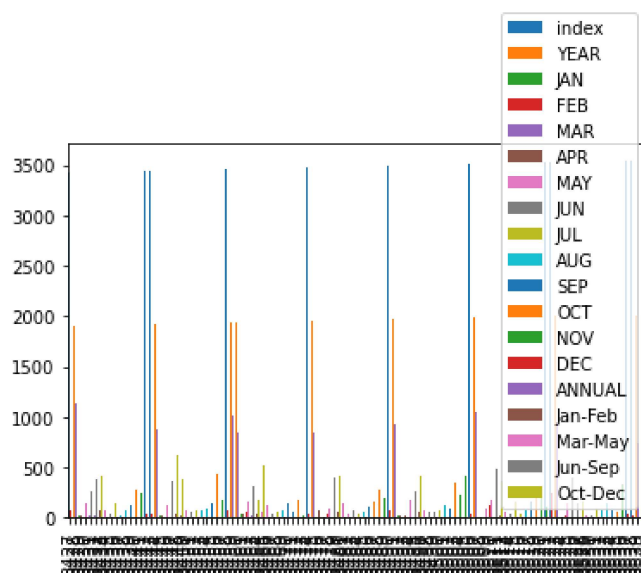
Line chart

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

Out[6]: `<AxesSubplot:>`
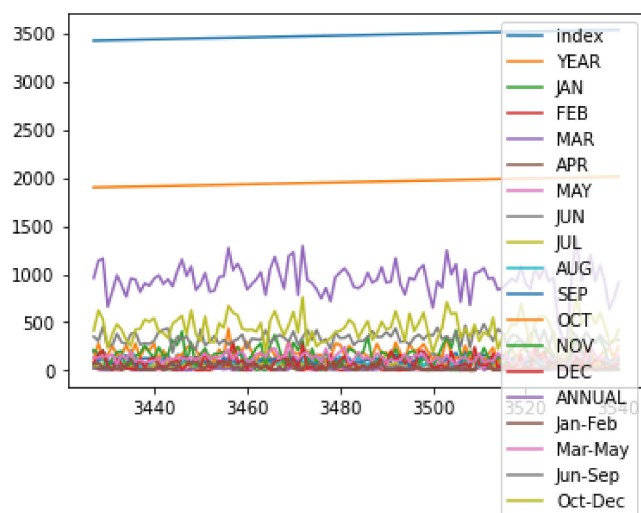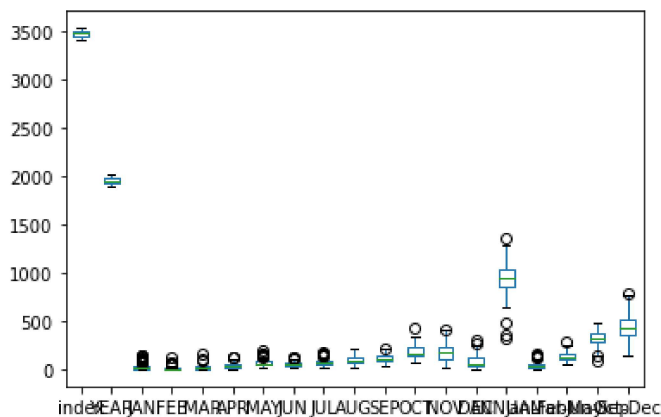

Bar chart

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

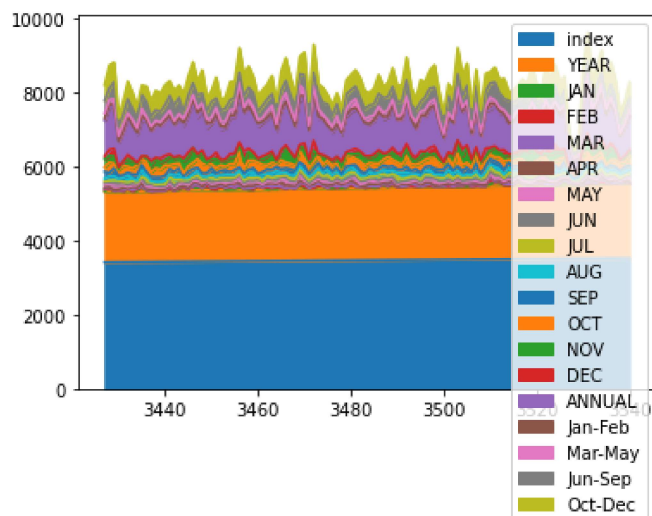Out[7]: `<AxesSubplot:>`

Box chart

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

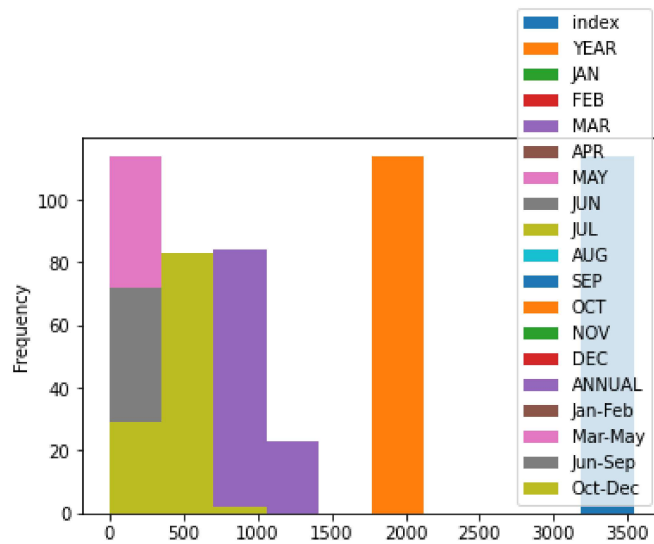Out[8]: `<AxesSubplot:>`

Area chart

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

Out[9]: `<AxesSubplot:>`

Histogram

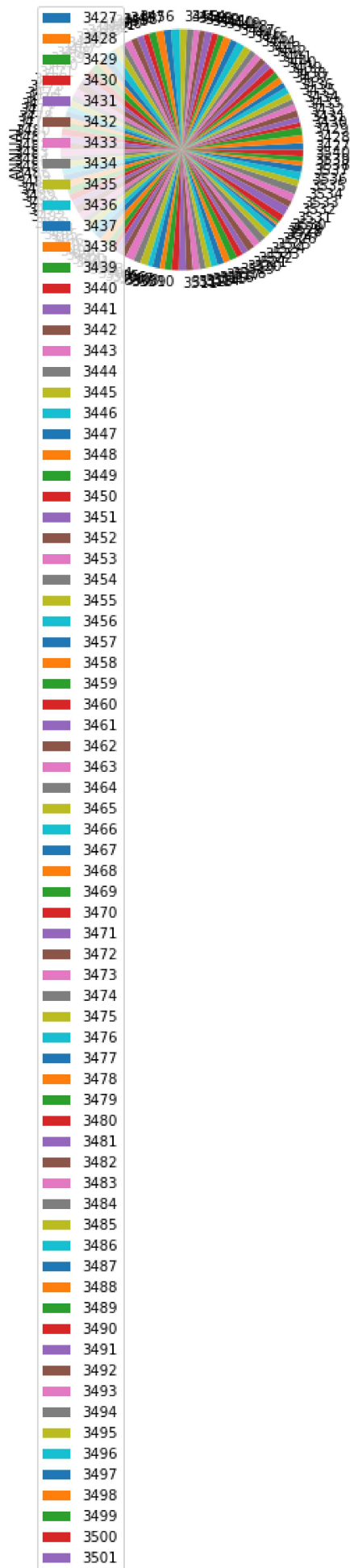In [10]: `df.plot.hist()`

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

pie chart

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

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

■	3502
■	3503
■	3504
■	3505
■	3506
■	3507
■	3508
■	3509
■	3510
■	3511
■	3512
■	3513
■	3514
■	3515
■	3516
■	3517
■	3518
■	3519
■	3520
■	3521
■	3522
■	3523
■	3524
■	3525
■	3526
■	3527
■	3528
■	3529
■	3530
■	3531
■	3532
■	3533
■	3534
■	3535
■	3536
■	3537
■	3538
■	3539
■	3540
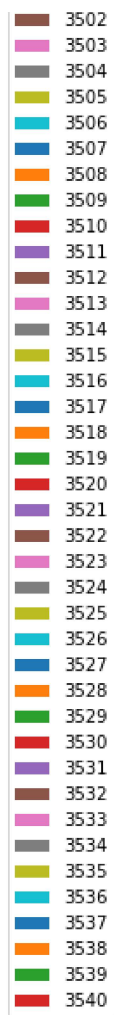
Scatter chart
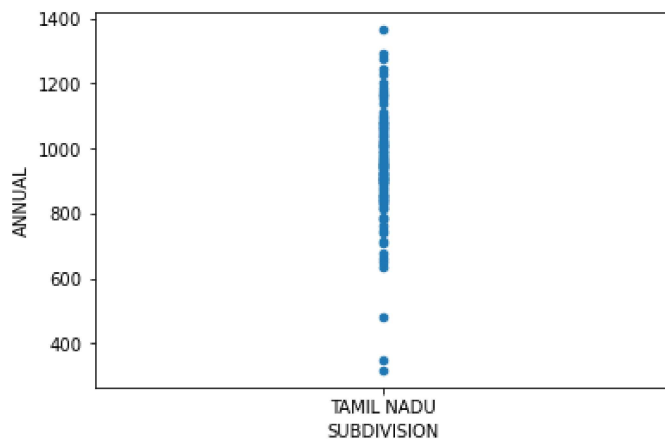
In [12]:
```
df.plot.scatter(y='ANNUAL',x='SUBDIVISION')
```

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	3483.500000	1957.500000	23.955263	13.520175	19.456140	44.435965	69.548246	51.965789	71.558772
std	33.052988	33.052988	32.362928	19.559106	22.487869	27.387860	31.662183	19.196210	31.086328
min	3427.000000	1901.000000	0.100000	0.000000	0.000000	5.500000	19.800000	21.200000	11.800000
25%	3455.250000	1929.250000	2.925000	1.200000	5.075000	23.625000	49.825000	38.625000	49.225000
50%	3483.500000	1957.500000	10.350000	5.550000	11.900000	36.500000	61.000000	49.800000	65.400000
75%	3511.750000	1985.750000	29.500000	18.075000	26.700000	58.450000	81.625000	62.500000	86.400000
max	3540.000000	2014.000000	141.200000	131.300000	164.700000	132.100000	204.400000	128.100000	174.700000

EDA AND VISUALIZATION

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

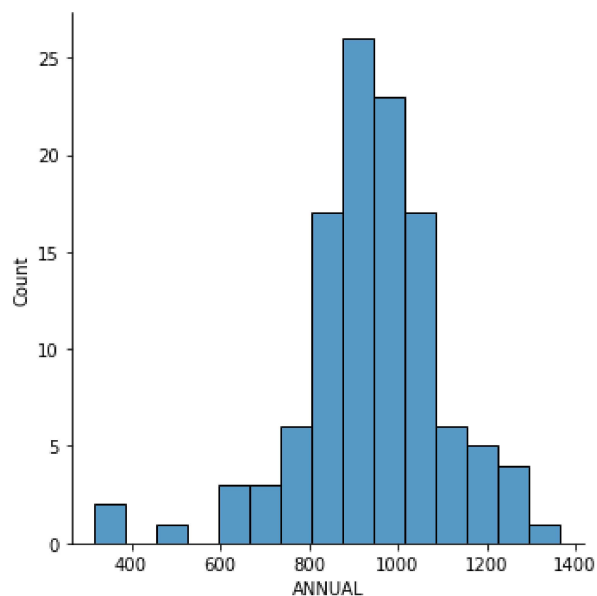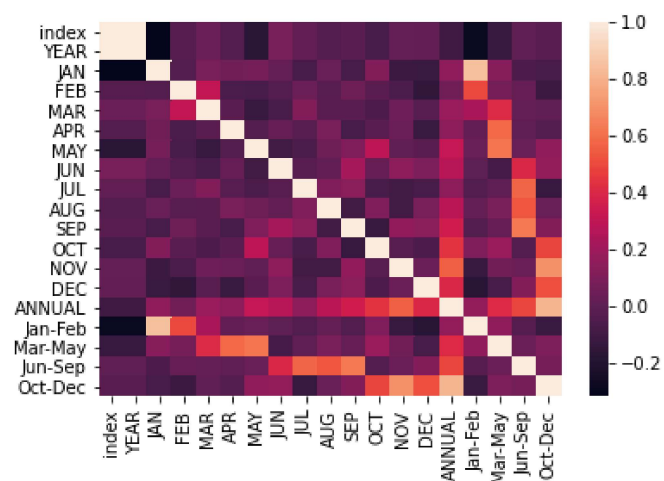Out[14]: `<seaborn.axisgrid.PairGrid at 0x1762494c6d0>`
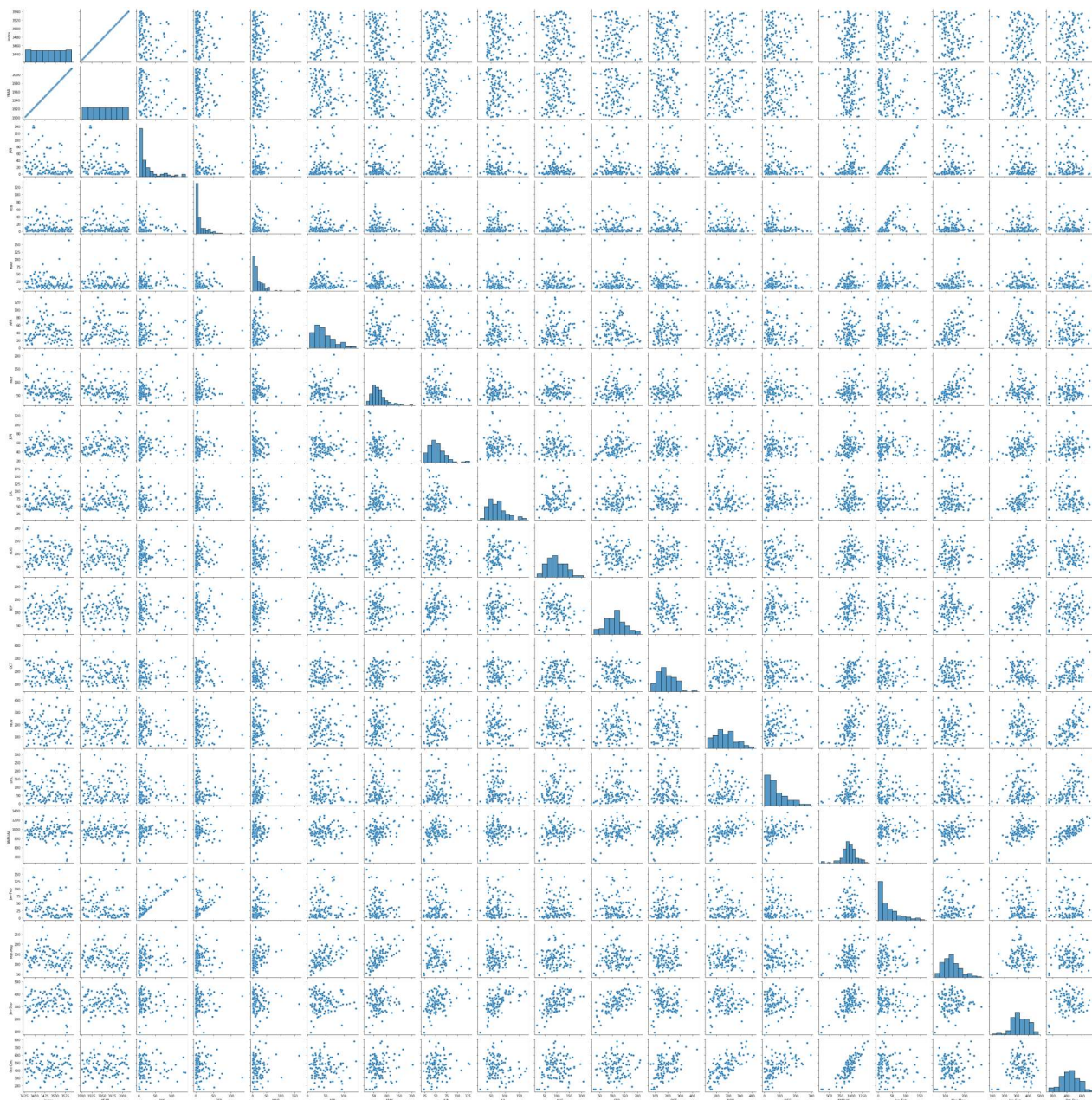
In [15]: `sns.displot(df['ANNUAL'])`

Out[15]: `<seaborn.axisgrid.FacetGrid at 0x17630dc19d0>`

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

Out[16]: `<AxesSubplot:>`

In [ ]: