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 1901-2(
         df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OC
3312	3312	RAYALSEEMA	1901	7.0	50.2	0.0	12.1	38.9	53.0	73.4	60.3	109.0	81.(
3313	3313	RAYALSEEMA	1902	10.0	0.2	1.7	11.0	36.8	73.6	41.3	148.3	181.7	188.!
3314	3314	RAYALSEEMA	1903	30.0	0.1	0.0	3.6	80.5	67.5	127.5	140.6	219.7	95.:
3315	3315	RAYALSEEMA	1904	14.8	0.0	1.7	7.1	58.8	39.8	75.1	19.4	84.7	111.!
3316	3316	RAYALSEEMA	1905	6.5	6.8	17.0	18.3	44.2	66.1	50.9	219.3	36.5	180.:
...	...	...	...	...	...	...	...	...	...	...	...	...	..
3421	3421	RAYALSEEMA	2010	4.7	0.8	0.6	13.0	65.4	108.3	187.6	155.0	122.9	83.4
3422	3422	RAYALSEEMA	2011	0.8	12.1	0.0	34.6	33.0	44.5	128.9	163.6	71.2	107.!
3423	3423	RAYALSEEMA	2012	2.7	0.0	2.5	32.7	38.8	47.0	139.7	120.0	69.5	113.7
3424	3424	RAYALSEEMA	2013	1.3	30.6	11.5	26.8	38.9	73.8	95.7	110.3	163.2	169.:
3425	3425	RAYALSEEMA	2014	0.2	0.7	12.5	5.1	46.7	66.3	68.7	115.1	81.4	104.(

114 rows × 20 columns

Data Cleaning and Preprocessing

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

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OC˙
3312	3312	RAYALSEEMA	1901	7.0	50.2	0.0	12.1	38.9	53.0	73.4	60.3	109.0	81.6
3313	3313	RAYALSEEMA	1902	10.0	0.2	1.7	11.0	36.8	73.6	41.3	148.3	181.7	188.5
3314	3314	RAYALSEEMA	1903	30.0	0.1	0.0	3.6	80.5	67.5	127.5	140.6	219.7	95.3
3315	3315	RAYALSEEMA	1904	14.8	0.0	1.7	7.1	58.8	39.8	75.1	19.4	84.7	111.5
3316	3316	RAYALSEEMA	1905	6.5	6.8	17.0	18.3	44.2	66.1	50.9	219.3	36.5	180.2
...	...	...	...	...	...	...	...	...	...	...	...	...	..
3421	3421	RAYALSEEMA	2010	4.7	0.8	0.6	13.0	65.4	108.3	187.6	155.0	122.9	83.4
3422	3422	RAYALSEEMA	2011	0.8	12.1	0.0	34.6	33.0	44.5	128.9	163.6	71.2	107.5
3423	3423	RAYALSEEMA	2012	2.7	0.0	2.5	32.7	38.8	47.0	139.7	120.0	69.5	113.7
3424	3424	RAYALSEEMA	2013	1.3	30.6	11.5	26.8	38.9	73.8	95.7	110.3	163.2	169.3
3425	3425	RAYALSEEMA	2014	0.2	0.7	12.5	5.1	46.7	66.3	68.7	115.1	81.4	104.6

114 rows × 20 columns

In [4]:
```python
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]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 114 entries, 3312 to 3425
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
```
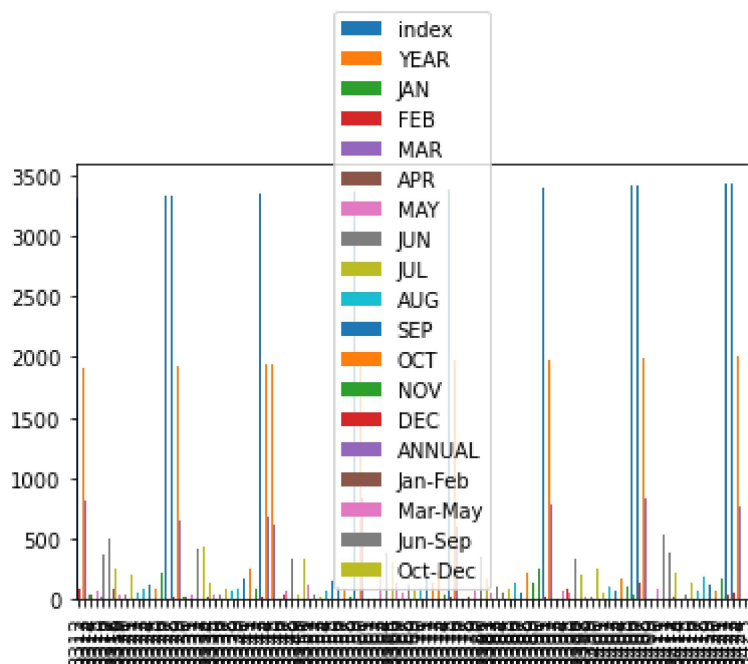
Line chart

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

Out[6]: <AxesSubplot:>
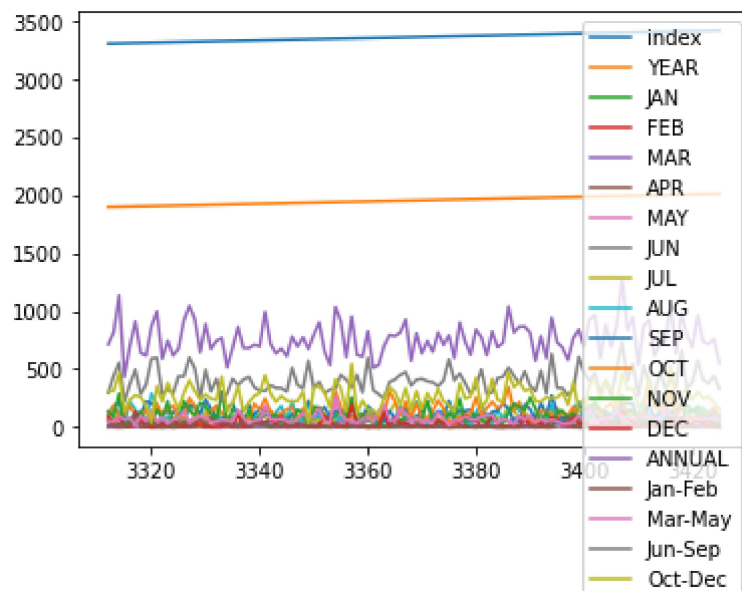
Bar chart

In [7]: `df.plot.bar()`
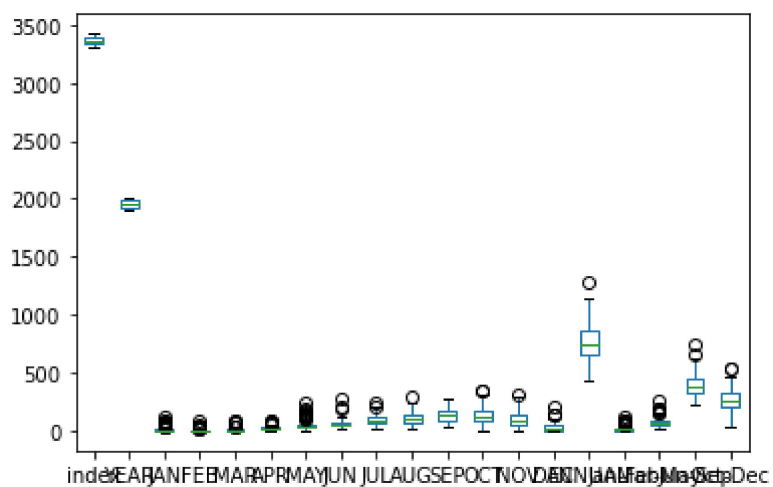
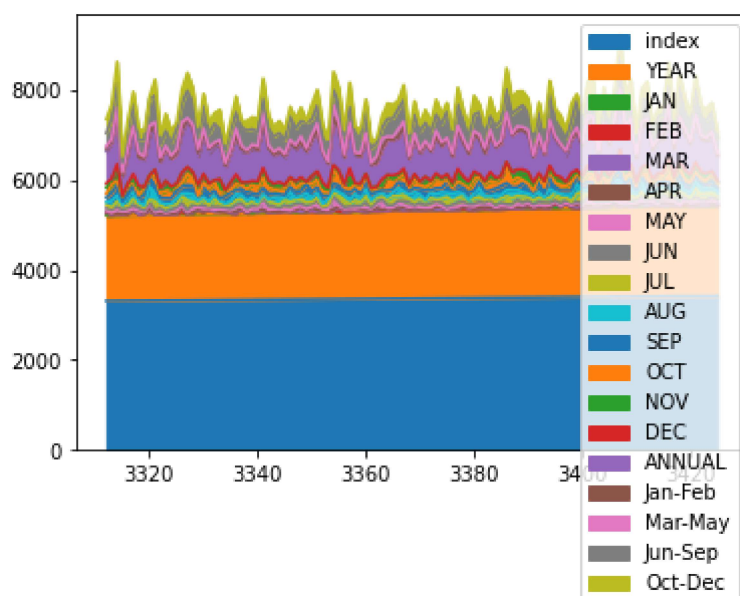Out[7]: `<AxesSubplot:>`

Box chart

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

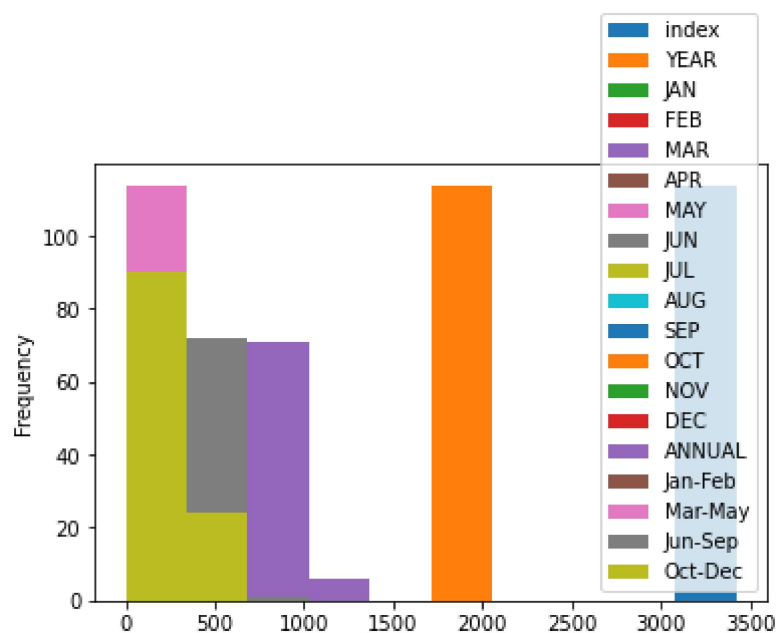Out[8]: `<AxesSubplot:>`

Area chart

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

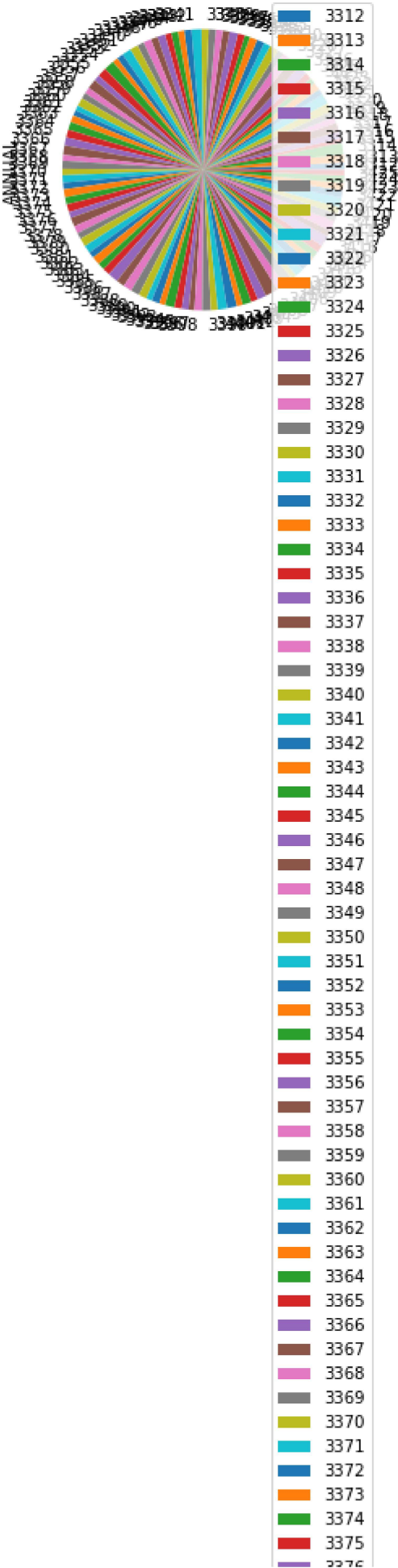Out[9]: `<AxesSubplot:>`

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'>`

3376
3377
3378
3379
3380
3381
3382
3383
3384
3385
3386
3387
3388
3389
3390
3391
3392
3393
3394
3395
3396
3397
3398
3399
3400
3401
3402
3403
3404
3405
3406
3407
3408
3409
3410
3411
3412
3413
3414
3415
3416
3417
3418
3419
3420
3421
3422
3423
3424
3425
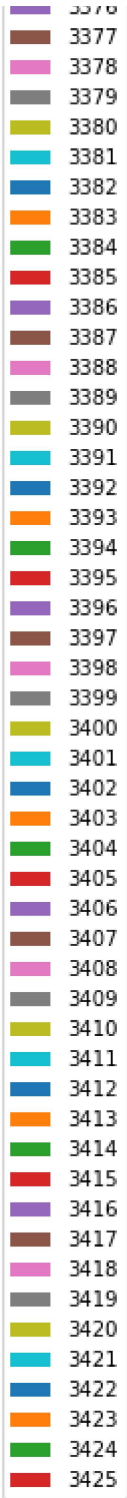
Scatter chart

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

Out[12]: `<AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>`
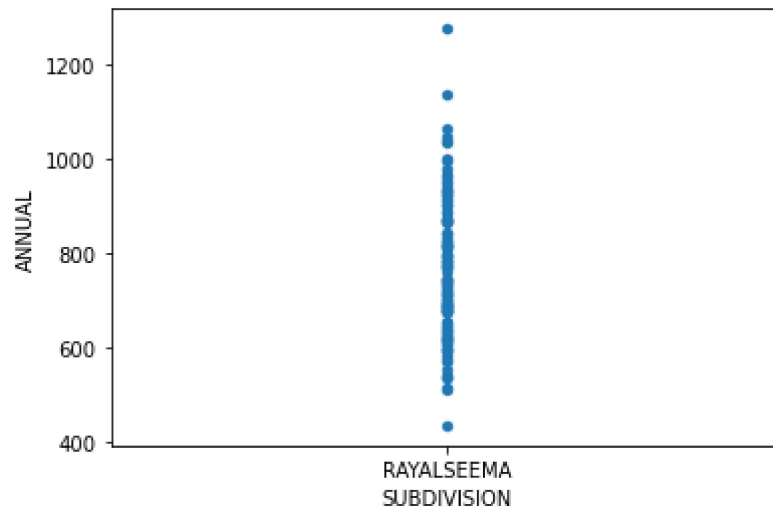
In [13]: `df.describe()`

Out[13]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	
count	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114
mean	3368.500000	1957.500000	9.937719	5.729825	8.029825	19.338596	50.570175	64
std	33.052988	33.052988	19.191399	11.691196	13.762929	16.911265	37.721412	34
min	3312.000000	1901.000000	0.000000	0.000000	0.000000	0.700000	4.100000	23
25%	3340.250000	1929.250000	0.200000	0.000000	0.250000	8.175000	29.075000	44
50%	3368.500000	1957.500000	2.000000	1.000000	3.950000	12.350000	41.450000	57
75%	3396.750000	1985.750000	9.950000	5.800000	10.725000	26.750000	55.225000	75
max	3425.000000	2014.000000	115.300000	81.000000	86.900000	93.500000	239.800000	27(

EDA AND VISUALIZATION

In [14]: sns.pairplot(df)

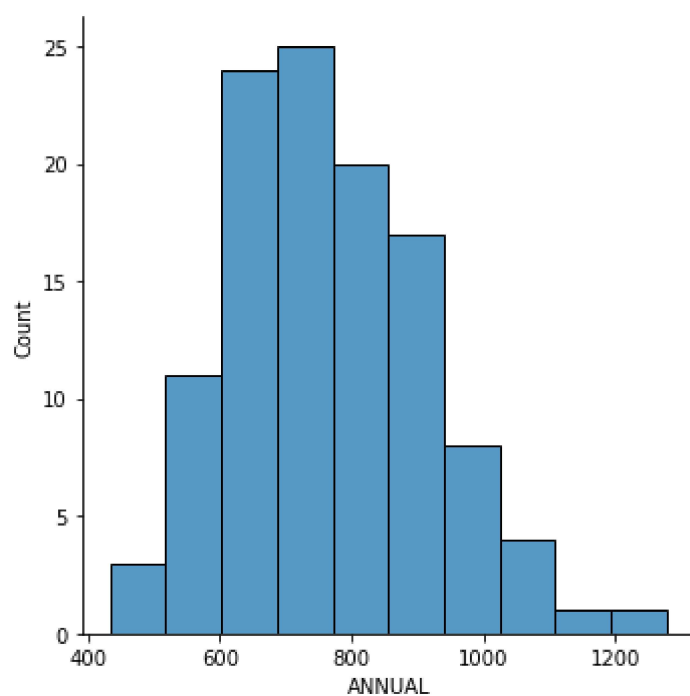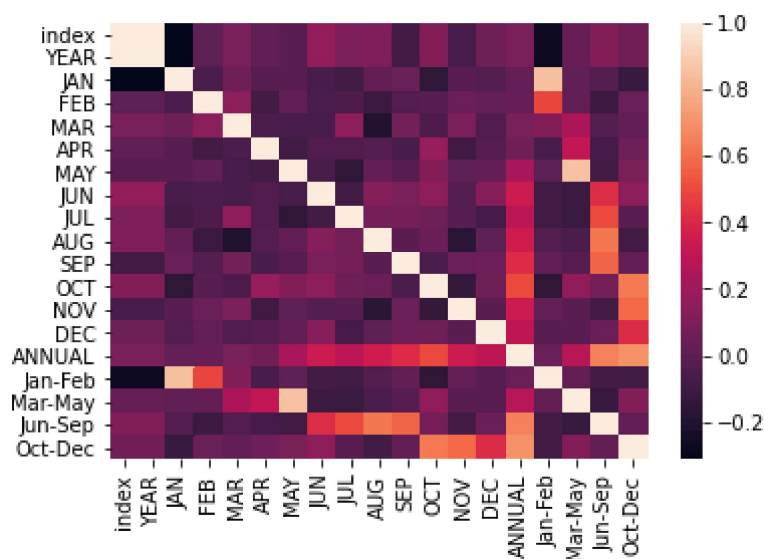Out[14]: <seaborn.axisgrid.PairGrid at 0x25d3c695670>
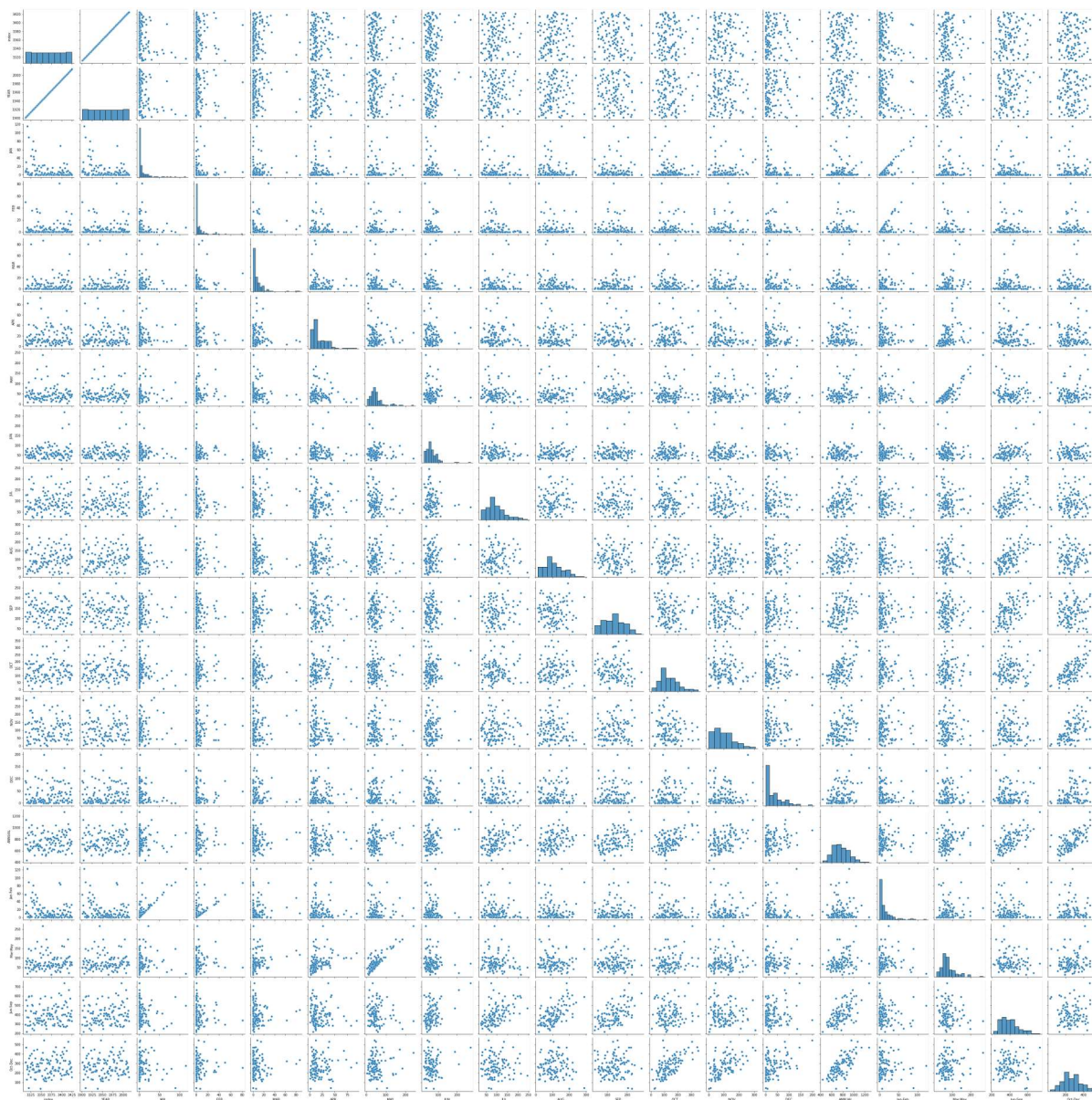
In [15]: `sns.displot(df['ANNUAL'])`

Out[15]: `<seaborn.axisgrid.FacetGrid at 0x25d48b35910>`

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

Out[16]: `<AxesSubplot:>`

In [ ]: