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")[3772:3886
        df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUA
3772	3772	SOUTH INTERIOR KARNATAKA	1901	4.9	31.8	3.0	32.7	109.6	106.0	210.0	109.2	140.8	170.1	72.5	12.3	1003
3773	3773	SOUTH INTERIOR KARNATAKA	1902	1.9	0.5	6.7	42.6	97.7	91.7	210.0	82.1	138.4	219.1	44.6	84.9	1020
3774	3774	SOUTH INTERIOR KARNATAKA	1903	0.3	0.0	1.1	11.6	125.1	129.7	284.4	155.7	197.1	154.2	186.6	24.1	1269
3775	3775	SOUTH INTERIOR KARNATAKA	1904	1.0	0.5	5.2	43.5	144.7	167.9	197.1	73.2	89.6	120.4	2.5	0.3	845
3776	3776	SOUTH INTERIOR KARNATAKA	1905	1.7	7.9	14.2	23.6	118.6	95.9	148.4	140.6	43.1	142.8	22.4	0.3	759
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
3881	3881	SOUTH INTERIOR KARNATAKA	2010	7.9	1.7	5.6	71.1	94.9	129.1	235.1	228.1	150.7	129.6	183.5	2.5	1239
3882	3882	SOUTH INTERIOR KARNATAKA	2011	2.1	12.4	12.4	80.2	83.5	177.1	202.4	199.5	111.2	144.8	56.7	5.0	1087
3883	3883	SOUTH INTERIOR KARNATAKA	2012	4.6	5.5	8.1	99.0	45.6	81.8	144.7	236.5	100.6	62.8	82.6	6.2	877
3884	3884	SOUTH INTERIOR KARNATAKA	2013	0.5	10.1	11.7	34.6	95.6	176.2	307.4	151.7	191.8	103.7	24.9	2.4	1110
3885	3885	SOUTH INTERIOR KARNATAKA	2014	0.4	2.4	17.7	46.7	130.5	106.8	271.6	254.6	161.6	152.9	20.2	18.7	1184

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	ANNUA
3772	3772	SOUTH INTERIOR KARNATAKA	1901	4.9	31.8	3.0	32.7	109.6	106.0	210.0	109.2	140.8	170.1	72.5	12.3	1003
3773	3773	SOUTH INTERIOR KARNATAKA	1902	1.9	0.5	6.7	42.6	97.7	91.7	210.0	82.1	138.4	219.1	44.6	84.9	1020
3774	3774	SOUTH INTERIOR KARNATAKA	1903	0.3	0.0	1.1	11.6	125.1	129.7	284.4	155.7	197.1	154.2	186.6	24.1	1269
3775	3775	SOUTH INTERIOR KARNATAKA	1904	1.0	0.5	5.2	43.5	144.7	167.9	197.1	73.2	89.6	120.4	2.5	0.3	845
3776	3776	SOUTH INTERIOR KARNATAKA	1905	1.7	7.9	14.2	23.6	118.6	95.9	148.4	140.6	43.1	142.8	22.4	0.3	759
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
3881	3881	SOUTH INTERIOR KARNATAKA	2010	7.9	1.7	5.6	71.1	94.9	129.1	235.1	228.1	150.7	129.6	183.5	2.5	1239
3882	3882	SOUTH INTERIOR KARNATAKA	2011	2.1	12.4	12.4	80.2	83.5	177.1	202.4	199.5	111.2	144.8	56.7	5.0	1087
3883	3883	SOUTH INTERIOR KARNATAKA	2012	4.6	5.5	8.1	99.0	45.6	81.8	144.7	236.5	100.6	62.8	82.6	6.2	877
3884	3884	SOUTH INTERIOR KARNATAKA	2013	0.5	10.1	11.7	34.6	95.6	176.2	307.4	151.7	191.8	103.7	24.9	2.4	1110
3885	3885	SOUTH INTERIOR KARNATAKA	2014	0.4	2.4	17.7	46.7	130.5	106.8	271.6	254.6	161.6	152.9	20.2	18.7	1184

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, 3772 to 3885
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()`

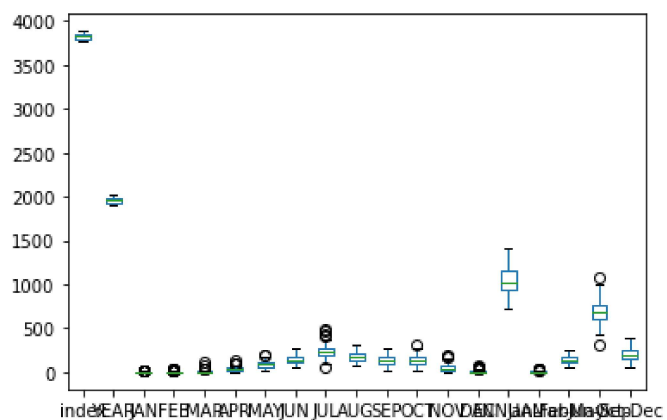Out[6]: `<AxesSubplot:>`
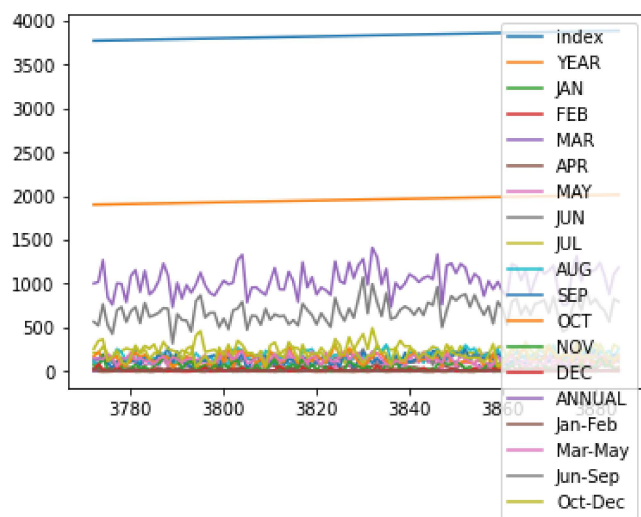
Bar chart

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

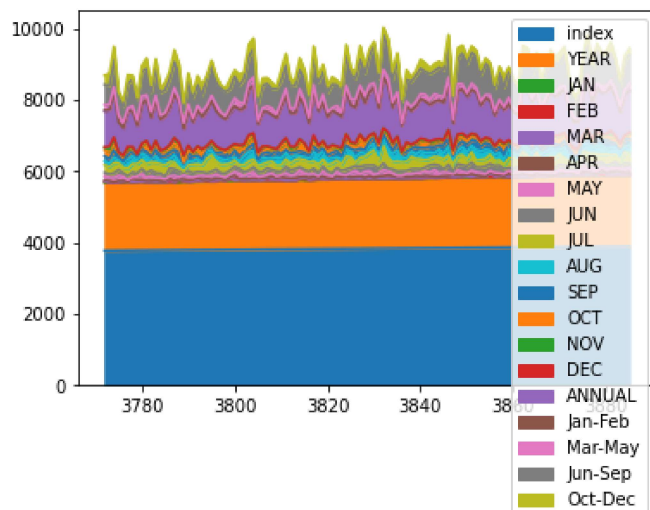Out[7]: `<AxesSubplot:>`
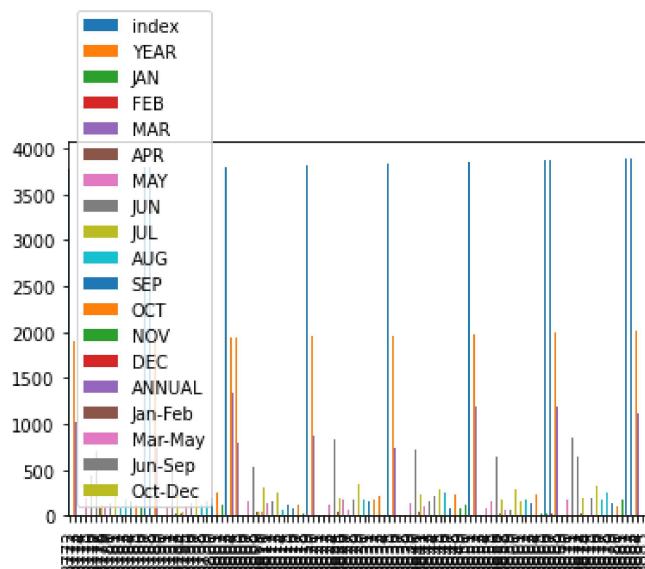
Box chart

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

Out[8]: `<AxesSubplot:>`

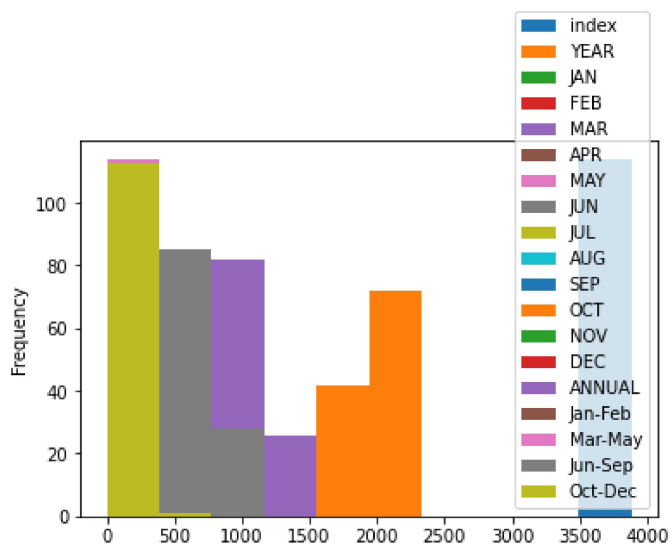# Area chart

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

Out[9]: `<AxesSubplot:>`

Histogram

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

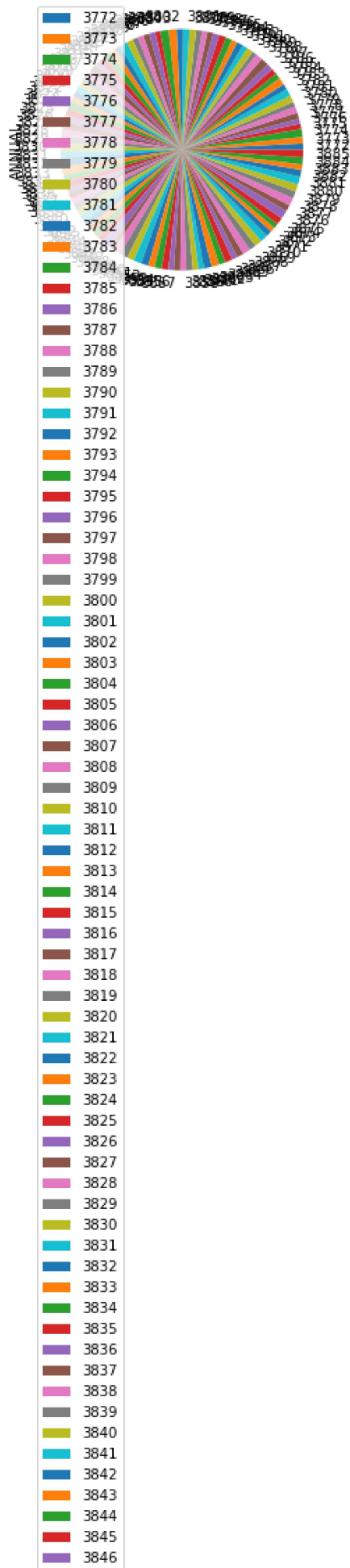Out[10]: `<AxesSubplot:ylabel='Frequency'>`

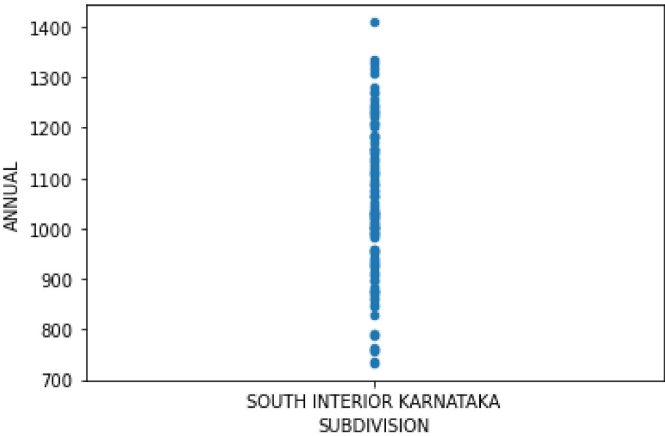# pie chart

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

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

3847
3848
3849
3850
3851
3852
3853
3854
3855
3856
3857
3858
3859
3860
3861
3862
3863
3864
3865
3866
3867
3868
3869
3870
3871
3872
3873
3874
3875
3876
3877
3878
3879
3880
3881
3882
3883
3884
3885

Scatter chart

In [12]:
```python
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	3828.500000	1957.500000	2.939474	4.198246	9.354386	41.945614	91.808772	140.739474	232.406140
std	33.052988	33.052988	4.828577	7.584579	13.946599	21.862216	38.259604	45.683927	70.903588
min	3772.000000	1901.000000	0.000000	0.000000	0.000000	4.800000	9.600000	64.000000	47.500000
25%	3800.250000	1929.250000	0.100000	0.100000	2.300000	25.425000	62.975000	106.800000	188.725000
50%	3828.500000	1957.500000	0.800000	1.200000	5.100000	39.900000	90.350000	135.800000	227.200000
75%	3856.750000	1985.750000	4.100000	4.600000	12.475000	50.850000	112.425000	173.225000	268.400000
max	3885.000000	2014.000000	24.400000	44.300000	108.900000	127.700000	190.500000	269.400000	492.700000

EDA AND VISUALIZATION

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

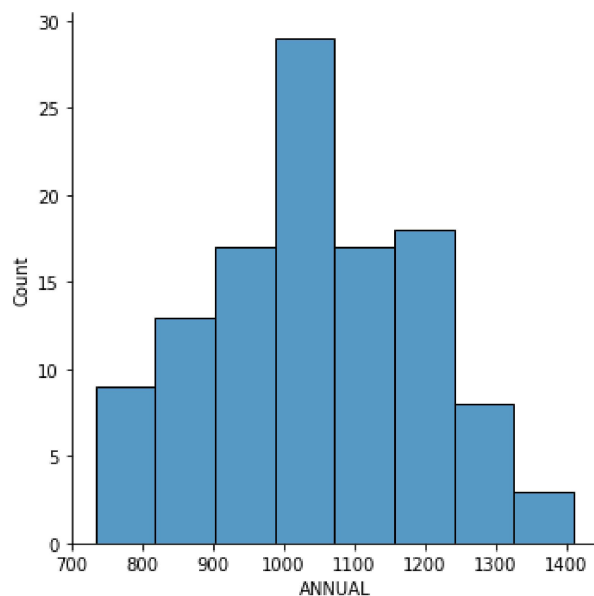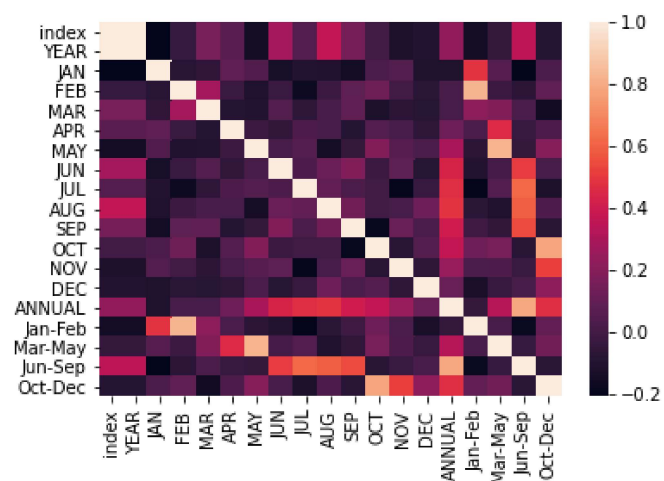Out[14]: `<seaborn.axisgrid.PairGrid at 0x25b4b783d30>`
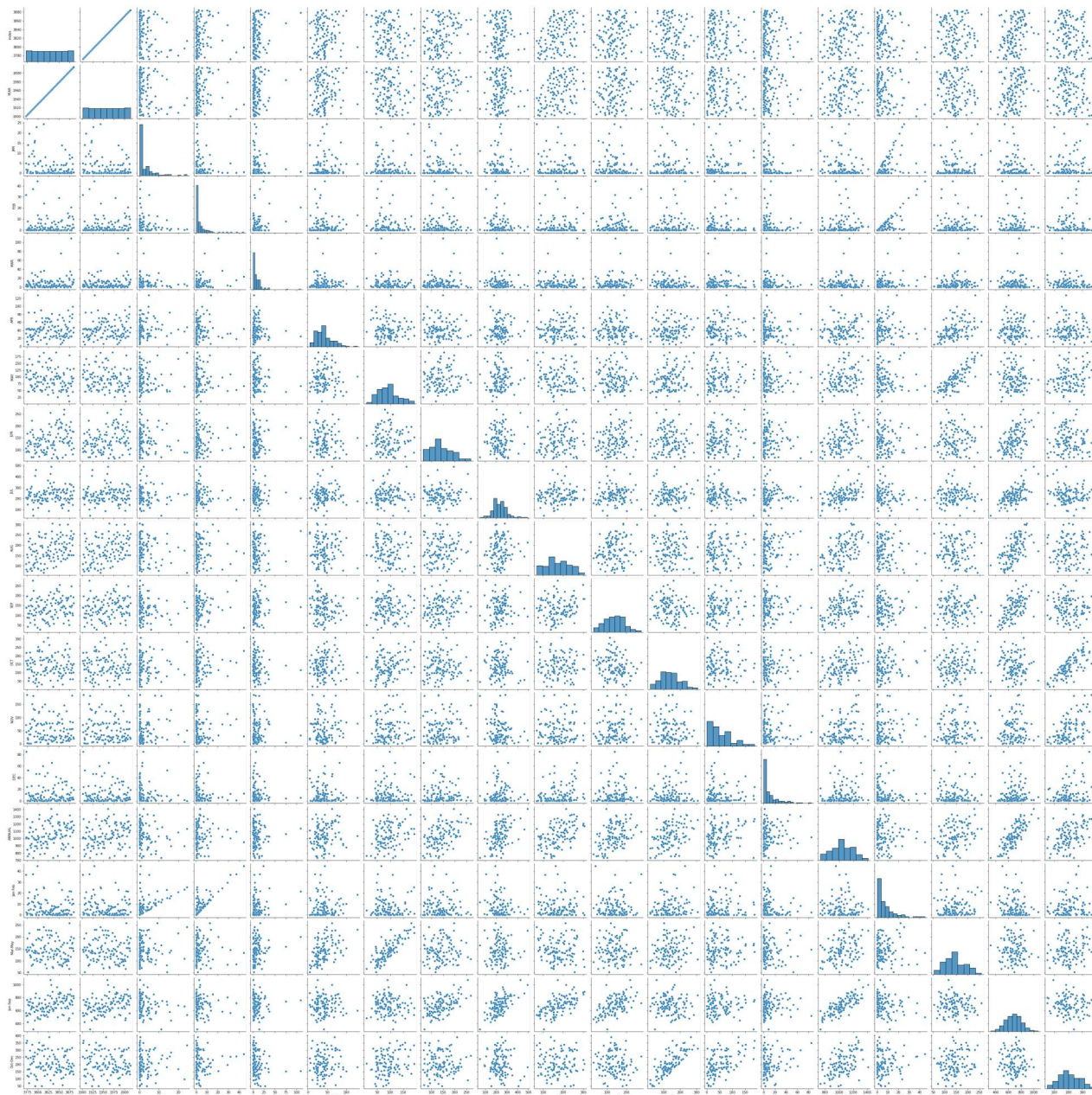
In [15]: `sns.displot(df['ANNUAL'])`

Out[15]: `<seaborn.axisgrid.FacetGrid at 0x25b5a383eb0>`

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

Out[16]: `<AxesSubplot:>`

In [ ]: