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")[3197:3311
         df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
3197	3197	TELANGANA	1901	6.9	41.8	7.8	45.2	22.0	123.6	237.8	177.2	77.7	75.5	12.2	0.0	827.7
3198	3198	TELANGANA	1902	0.0	0.0	0.2	10.7	7.3	52.4	146.3	142.8	190.5	41.7	31.2	7.3	630.4
3199	3199	TELANGANA	1903	12.9	4.6	0.0	9.9	40.7	99.2	505.2	246.7	191.9	155.8	15.5	1.1	1283.4
3200	3200	TELANGANA	1904	0.0	0.0	10.8	0.8	14.7	104.2	139.5	50.0	162.3	44.4	0.0	0.0	526.7
3201	3201	TELANGANA	1905	0.0	4.3	12.8	27.6	32.2	129.5	82.4	237.3	179.1	19.6	0.0	0.0	724.9
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..
3306	3306	TELANGANA	2010	10.3	5.3	1.5	5.6	24.9	127.0	395.5	308.1	249.8	98.7	40.5	9.3	1276.4
3307	3307	TELANGANA	2011	0.0	11.9	2.6	25.6	9.3	83.9	268.2	225.9	107.6	13.9	4.2	0.0	753.1
3308	3308	TELANGANA	2012	6.7	0.0	0.2	14.0	8.4	124.4	300.3	229.9	202.4	83.6	38.7	0.0	1008.6
3309	3309	TELANGANA	2013	2.4	29.0	0.2	24.4	8.5	213.4	453.8	230.6	161.4	205.9	16.4	2.7	1348.7
3310	3310	TELANGANA	2014	0.2	2.9	58.3	10.3	73.3	62.3	146.0	205.2	146.8	29.6	10.8	0.7	746.4

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
3197	3197	TELANGANA	1901	6.9	41.8	7.8	45.2	22.0	123.6	237.8	177.2	77.7	75.5	12.2	0.0	827.7
3198	3198	TELANGANA	1902	0.0	0.0	0.2	10.7	7.3	52.4	146.3	142.8	190.5	41.7	31.2	7.3	630.4
3199	3199	TELANGANA	1903	12.9	4.6	0.0	9.9	40.7	99.2	505.2	246.7	191.9	155.8	15.5	1.1	1283.4
3200	3200	TELANGANA	1904	0.0	0.0	10.8	0.8	14.7	104.2	139.5	50.0	162.3	44.4	0.0	0.0	526.7
3201	3201	TELANGANA	1905	0.0	4.3	12.8	27.6	32.2	129.5	82.4	237.3	179.1	19.6	0.0	0.0	724.9
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..
3306	3306	TELANGANA	2010	10.3	5.3	1.5	5.6	24.9	127.0	395.5	308.1	249.8	98.7	40.5	9.3	1276.4
3307	3307	TELANGANA	2011	0.0	11.9	2.6	25.6	9.3	83.9	268.2	225.9	107.6	13.9	4.2	0.0	753.1
3308	3308	TELANGANA	2012	6.7	0.0	0.2	14.0	8.4	124.4	300.3	229.9	202.4	83.6	38.7	0.0	1008.6
3309	3309	TELANGANA	2013	2.4	29.0	0.2	24.4	8.5	213.4	453.8	230.6	161.4	205.9	16.4	2.7	1348.7
3310	3310	TELANGANA	2014	0.2	2.9	58.3	10.3	73.3	62.3	146.0	205.2	146.8	29.6	10.8	0.7	746.4

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, 3197 to 3310
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
```
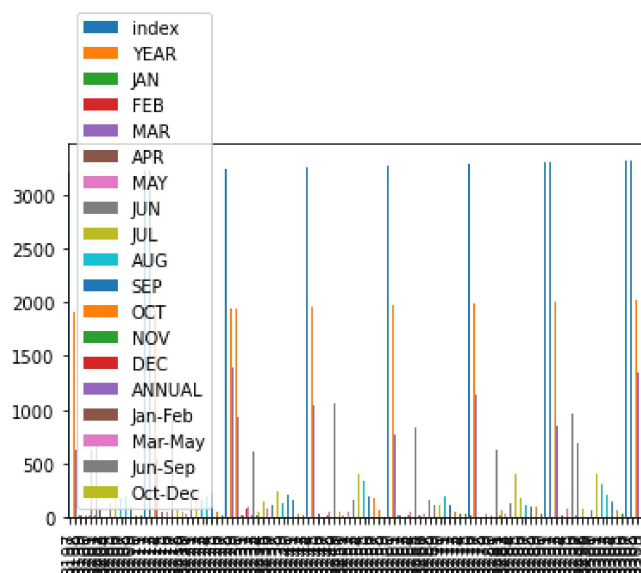
Line chart

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

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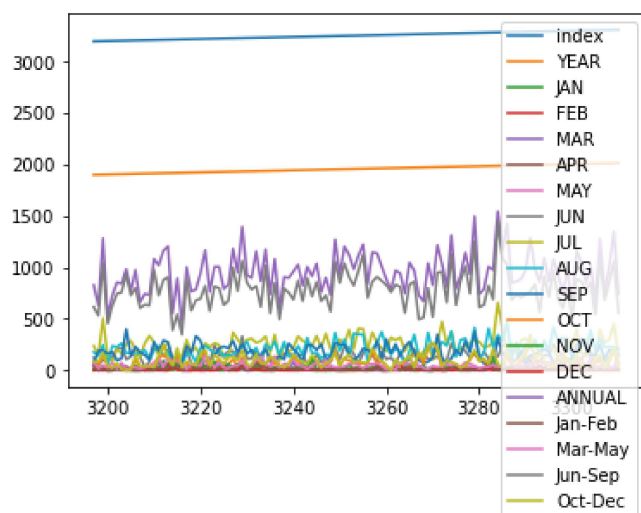
Bar chart

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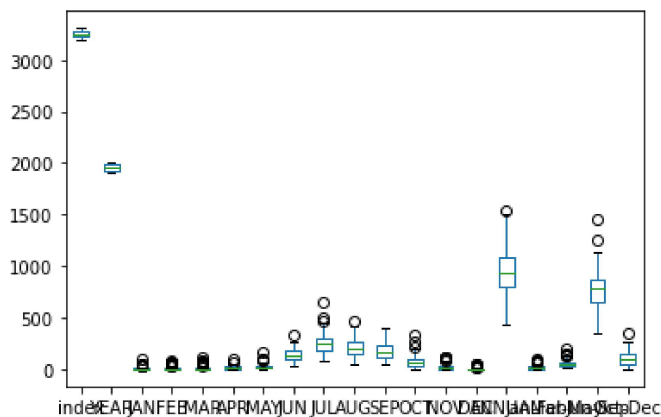
Out[7]: `<AxesSubplot:>`

Box chart

In [8]: `df.plot.box()`
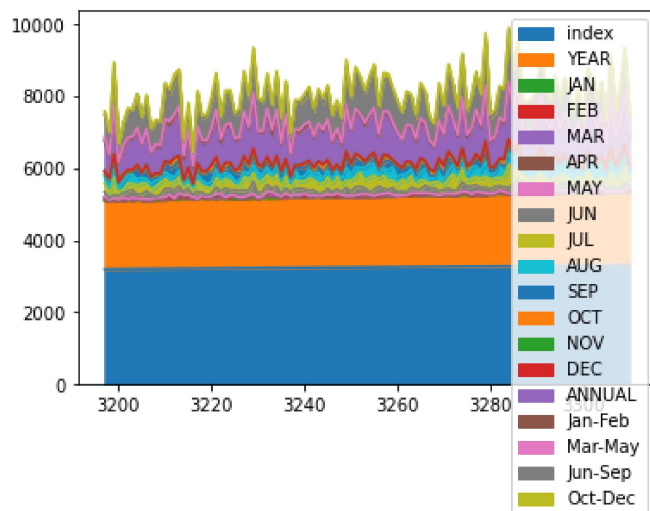
Out[8]: `<AxesSubplot:>`

Area chart

In [9]: `df.plot.area()`
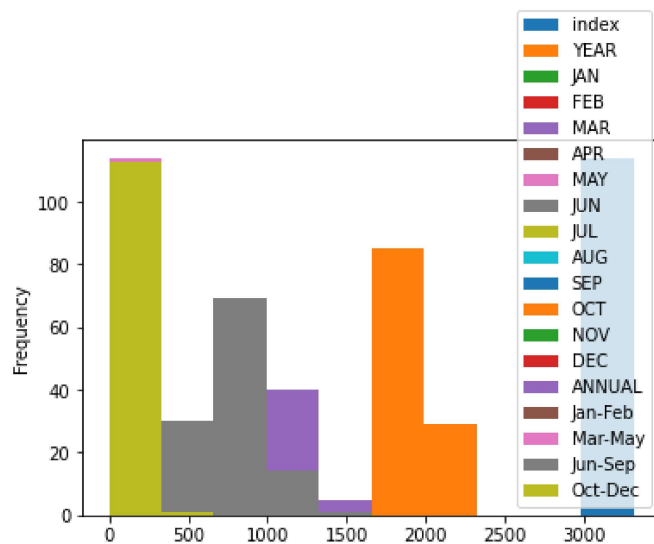
Out[9]: `<AxesSubplot:>`
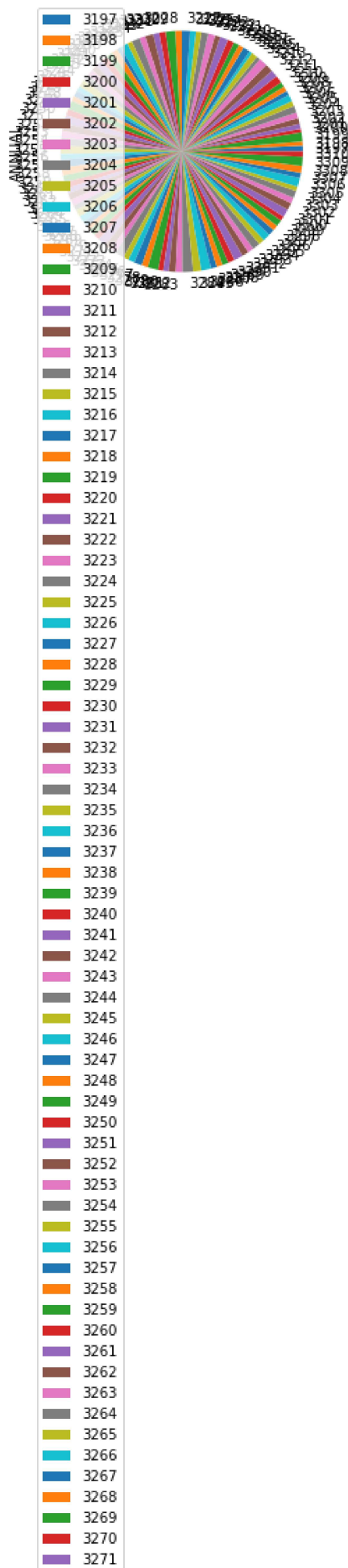

Histogram

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

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

pie chart

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

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

■	3272
■	3273
■	3274
■	3275
■	3276
■	3277
■	3278
■	3279
■	3280
■	3281
■	3282
■	3283
■	3284
■	3285
■	3286
■	3287
■	3288
■	3289
■	3290
■	3291
■	3292
■	3293
■	3294
■	3295
■	3296
■	3297
■	3298
■	3299
■	3300
■	3301
■	3302
■	3303
■	3304
■	3305
■	3306
■	3307
■	3308
■	3309
■	3310
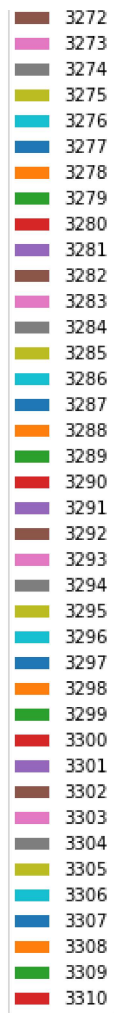
Scatter chart

In [12]:
```python
df.plot.scatter(y='ANNUAL',x='SUBDIVISION')
```
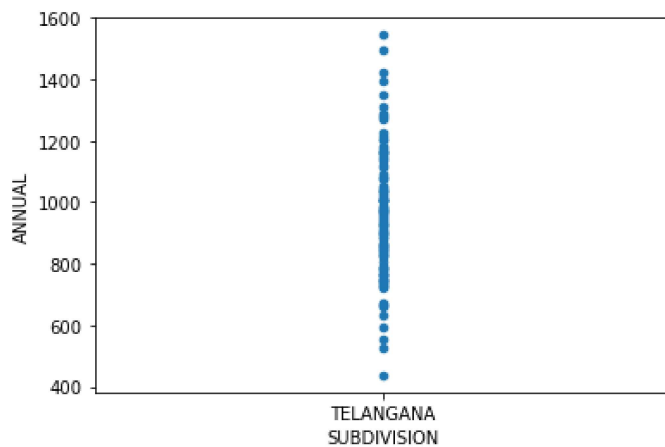
Out[12]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>

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	3253.500000	1957.500000	7.616667	9.773684	12.348246	17.768421	25.392105	141.031579	248.754386
std	33.052988	33.052988	13.911744	15.263596	18.645008	15.358975	23.746352	56.769201	95.526247
min	3197.000000	1901.000000	0.000000	0.000000	0.000000	0.100000	0.200000	28.500000	82.400000
25%	3225.250000	1929.250000	0.000000	0.000000	1.400000	6.925000	8.425000	103.450000	183.525000
50%	3253.500000	1957.500000	1.000000	3.750000	4.700000	14.000000	20.500000	133.450000	249.350000
75%	3281.750000	1985.750000	9.350000	14.000000	15.050000	24.625000	34.550000	180.475000	296.850000
max	3310.000000	2014.000000	98.700000	79.100000	108.600000	105.600000	159.800000	332.000000	656.200000

EDA AND VISUALIZATION

In [14]: sns.pairplot(df)

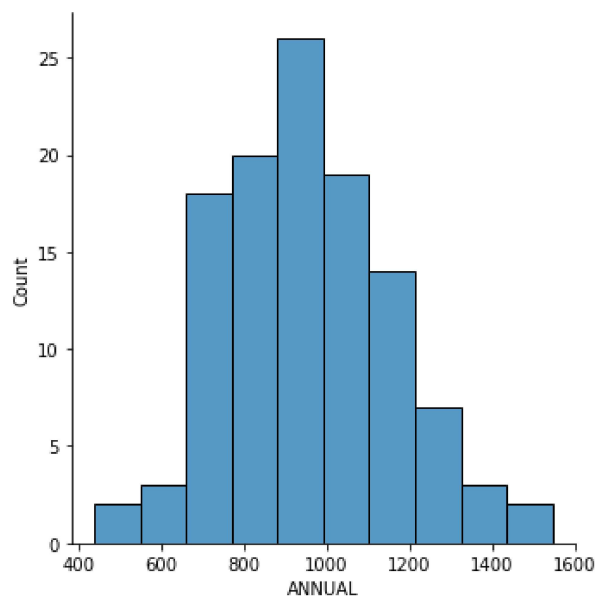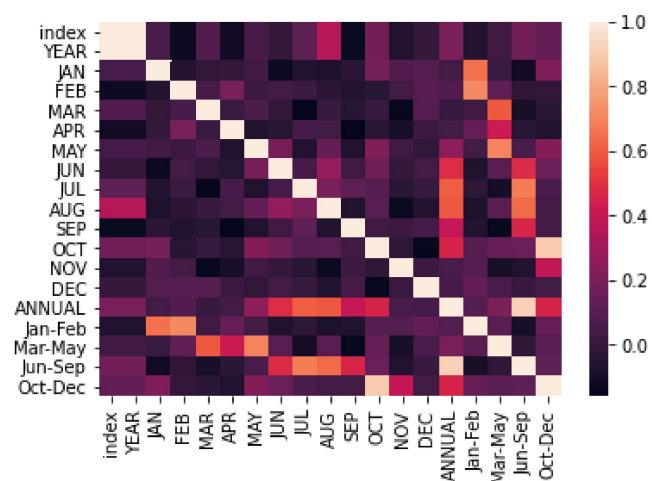Out[14]: <seaborn.axisgrid.PairGrid at 0x21d66ad1700>
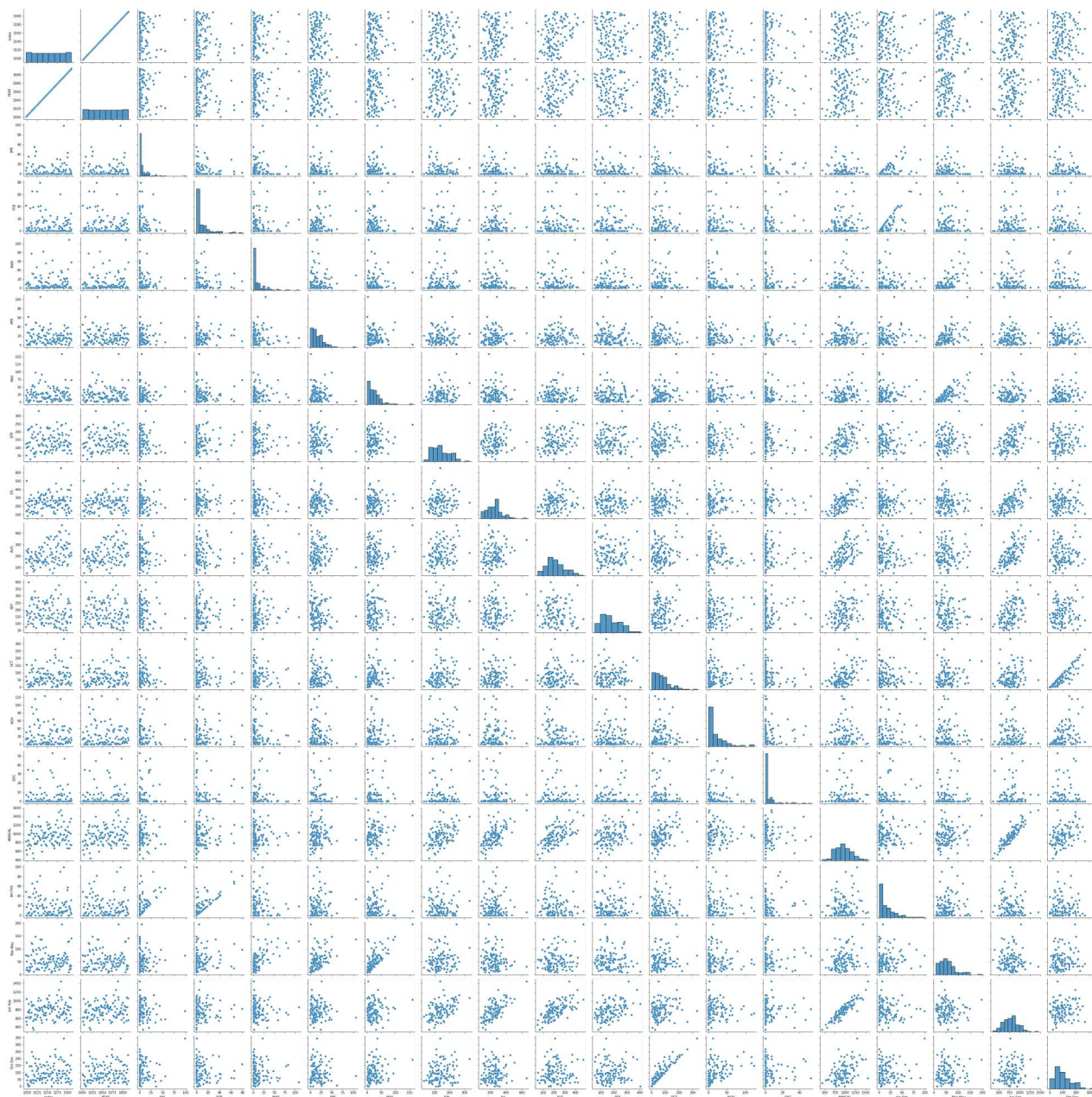
In [15]: `sns.displot(df['ANNUAL'])`

Out[15]: `<seaborn.axisgrid.FacetGrid at 0x21d75b35250>`

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

Out[16]: `<AxesSubplot:>`

In [ ]: