FINAL ASSESSMENT 2

```
In [1]:  #importing libraries
         import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         import seaborn as sns
```

```
In [2]:  #importing dataset
         data=pd.read_csv(r"C:\Users\user\Downloads\rainfall in india 1901-2015.csv")
         data
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
0	0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6
1	1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2
2	2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0
3	3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4
4	4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0
...	...	...	...	...	...	...	...	...	...	...	...	...
4111	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2
4112	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8
4113	4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.0
4114	4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.2
4115	4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	296.6	257.5	146.4	160.4

4116 rows × 20 columns

TAMIL NADU

In [3]:
```
df=data.iloc[3427:3542]
df
```

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
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
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
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
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
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
...	...	...	...	...	...	...	...	...	...	...	...	...	...
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
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
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
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
3541	3541	TAMIL NADU	2015	8.3	2.3	21.7	108.8	112.4	62.4	43.5	81.6	98.4	132.6

115 rows × 20 columns

Data Cleaning and Preprocessing

In [4]:
```
df.head()
```

Out[4]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
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
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
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
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
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

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

Out[5]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
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
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
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
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
3541	3541	TAMIL NADU	2015	8.3	2.3	21.7	108.8	112.4	62.4	43.5	81.6	98.4	132.6

In [6]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 115 entries, 3427 to 3541
Data columns (total 20 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   index        115 non-null     int64
 1   SUBDIVISION  115 non-null     object
 2   YEAR         115 non-null     int64
 3   JAN          115 non-null     float64
 4   FEB          115 non-null     float64
 5   MAR          115 non-null     float64
 6   APR          115 non-null     float64
 7   MAY          115 non-null     float64
 8   JUN          115 non-null     float64
 9   JUL          115 non-null     float64
 10  AUG          115 non-null     float64
 11  SEP          115 non-null     float64
 12  OCT          115 non-null     float64
 13  NOV          115 non-null     float64
 14  DEC          115 non-null     float64
 15  ANNUAL       115 non-null     float64
 16  Jan-Feb      115 non-null     float64
 17  Mar-May      115 non-null     float64
 18  Jun-Sep      115 non-null     float64
 19  Oct-Dec      115 non-null     float64
dtypes: float64(17), int64(2), object(1)
memory usage: 18.1+ KB
```

In [7]:
```python
#filling null values
df1=df.fillna(0)
df1
```

Out[7]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
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
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
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
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
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
...	...	...	...	...	...	...	...	...	...	...	...	...	...
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
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
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
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
3541	3541	TAMIL NADU	2015	8.3	2.3	21.7	108.8	112.4	62.4	43.5	81.6	98.4	132.6

115 rows × 20 columns

In [8]:
```python
df1.describe()
```

Out[8]:

|	index	YEAR	JAN	FEB	MAR	APR	MAY
count	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.
mean	3484.000000	1958.000000	23.819130	13.422609	19.475652	44.995652	69.920870	52.
std	33.341666	33.341666	32.253728	19.501219	22.389999	27.920223	31.775268	19.
min	3427.000000	1901.000000	0.100000	0.000000	0.000000	5.500000	19.800000	21.
25%	3455.500000	1929.500000	2.950000	1.200000	5.150000	23.650000	50.050000	38.
50%	3484.000000	1958.000000	10.000000	5.500000	11.900000	37.000000	61.100000	49.
75%	3512.500000	1986.500000	29.300000	18.050000	26.700000	59.300000	82.400000	62.
max	3541.000000	2015.000000	141.200000	131.300000	164.700000	132.100000	204.400000	128.

In [9]:
```python
df1.columns
```

Out[9]:
```
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')
```
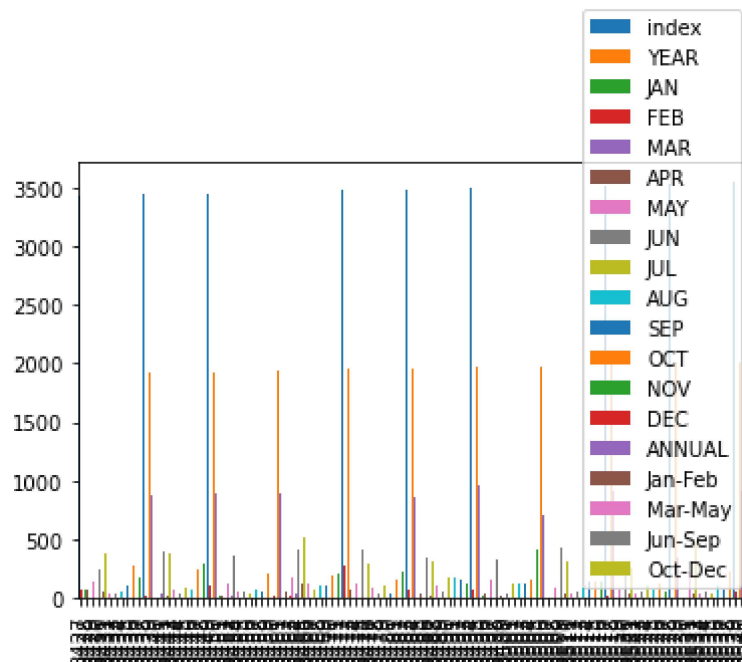
Data Visulaization
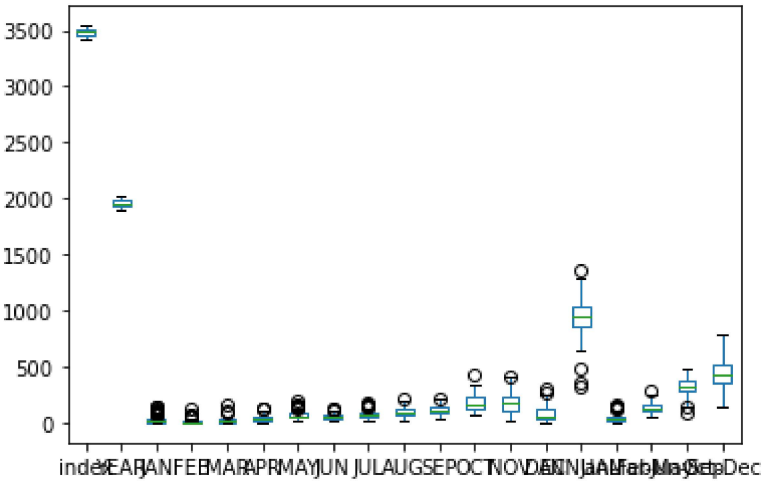
In [10]: `df1.plot.line()`

Out[10]: `<AxesSubplot:>`
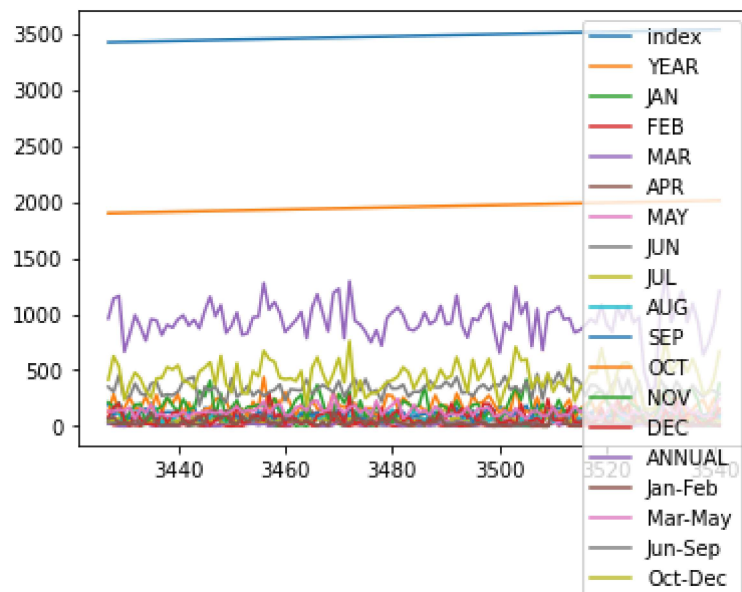
In [11]: `df1.plot.bar()`

Out[11]: `<AxesSubplot:>`
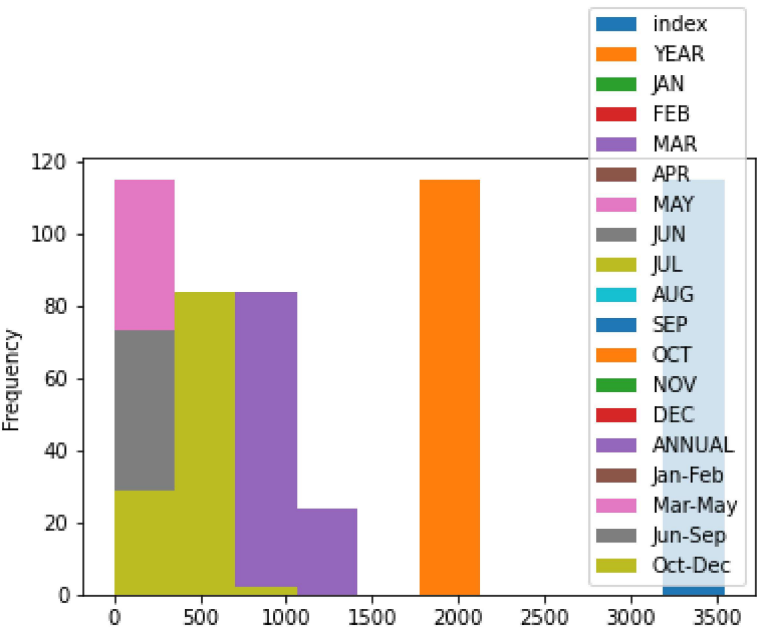
In [12]: `df1.plot.box()`

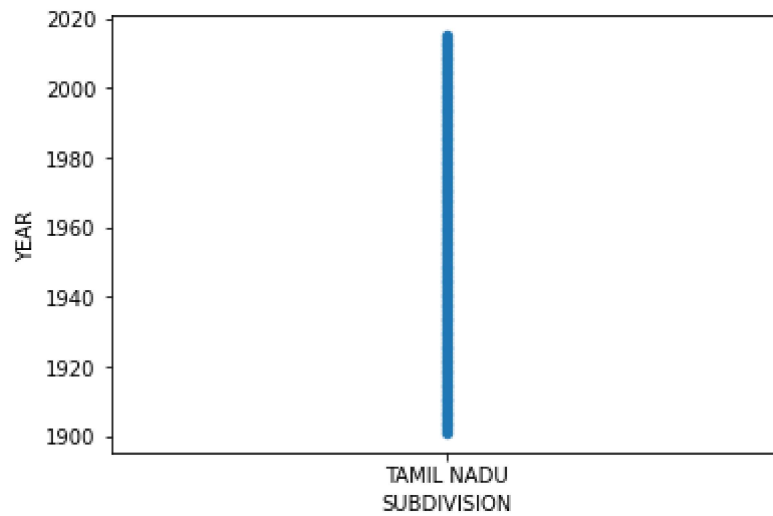Out[12]: `<AxesSubplot:>`

In [13]: `df1.plot.hist()`

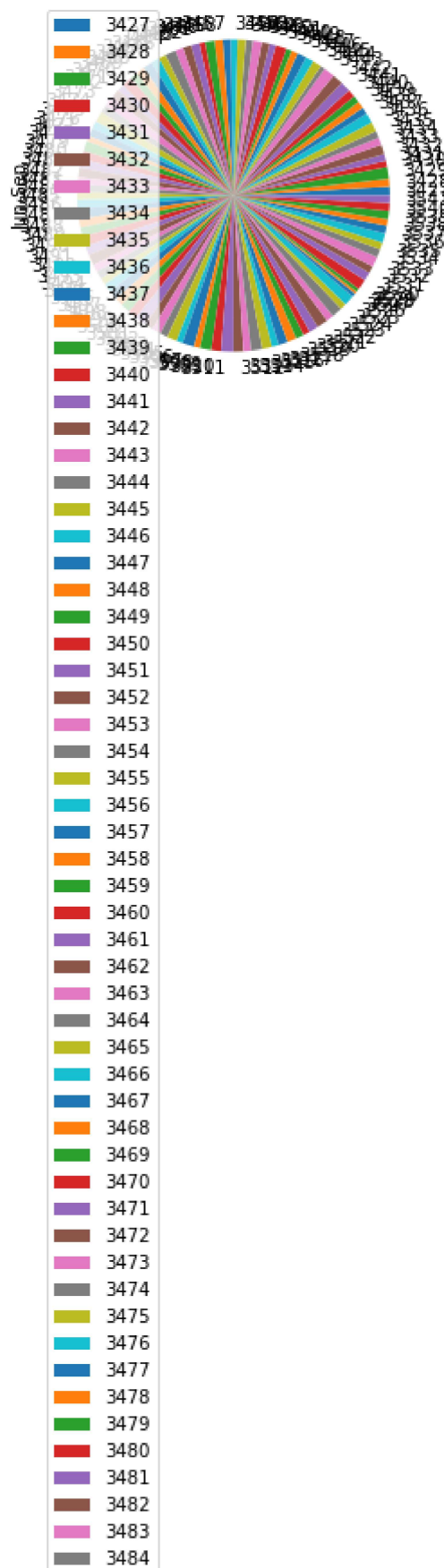Out[13]: `<AxesSubplot:ylabel='Frequency'>`

In [14]: `df1.plot.scatter(x="SUBDIVISION",y="YEAR")`

Out[14]: `<AxesSubplot:xlabel='SUBDIVISION', ylabel='YEAR'>`

In [15]:
```python
df2=df1[[ 'Jun-Sep']]
df2.plot.pie(subplots=True)
```

Out[15]: array([<AxesSubplot:ylabel='Jun-Sep'>], dtype=object)

3485
3486
3487
3488
3489
3490
3491
3492
3493
3494
3495
3496
3497
3498
3499
3500
3501
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
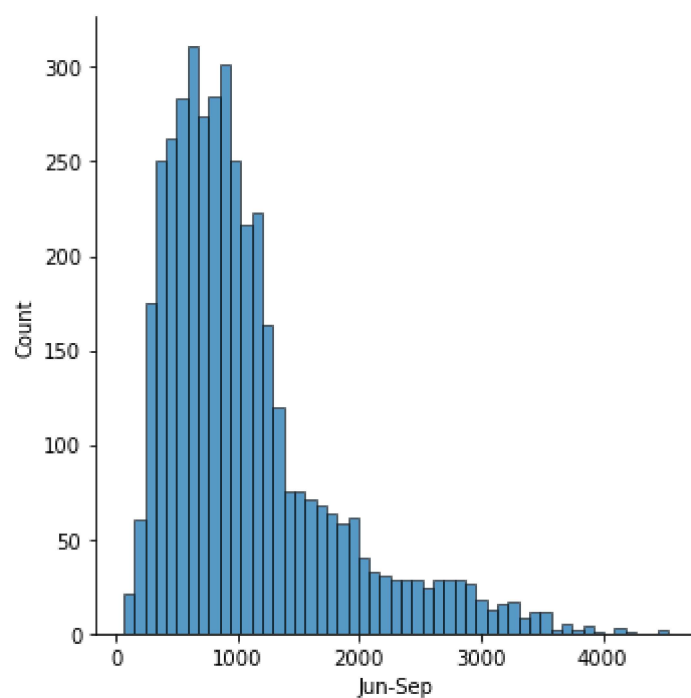3531
3532
3533
3534
3535
3536
3537
3538
3539
3540
3541

In [16]: `sns.pairplot(df1)`

Out[16]: `<seaborn.axisgrid.PairGrid at 0x15b6334fac0>`
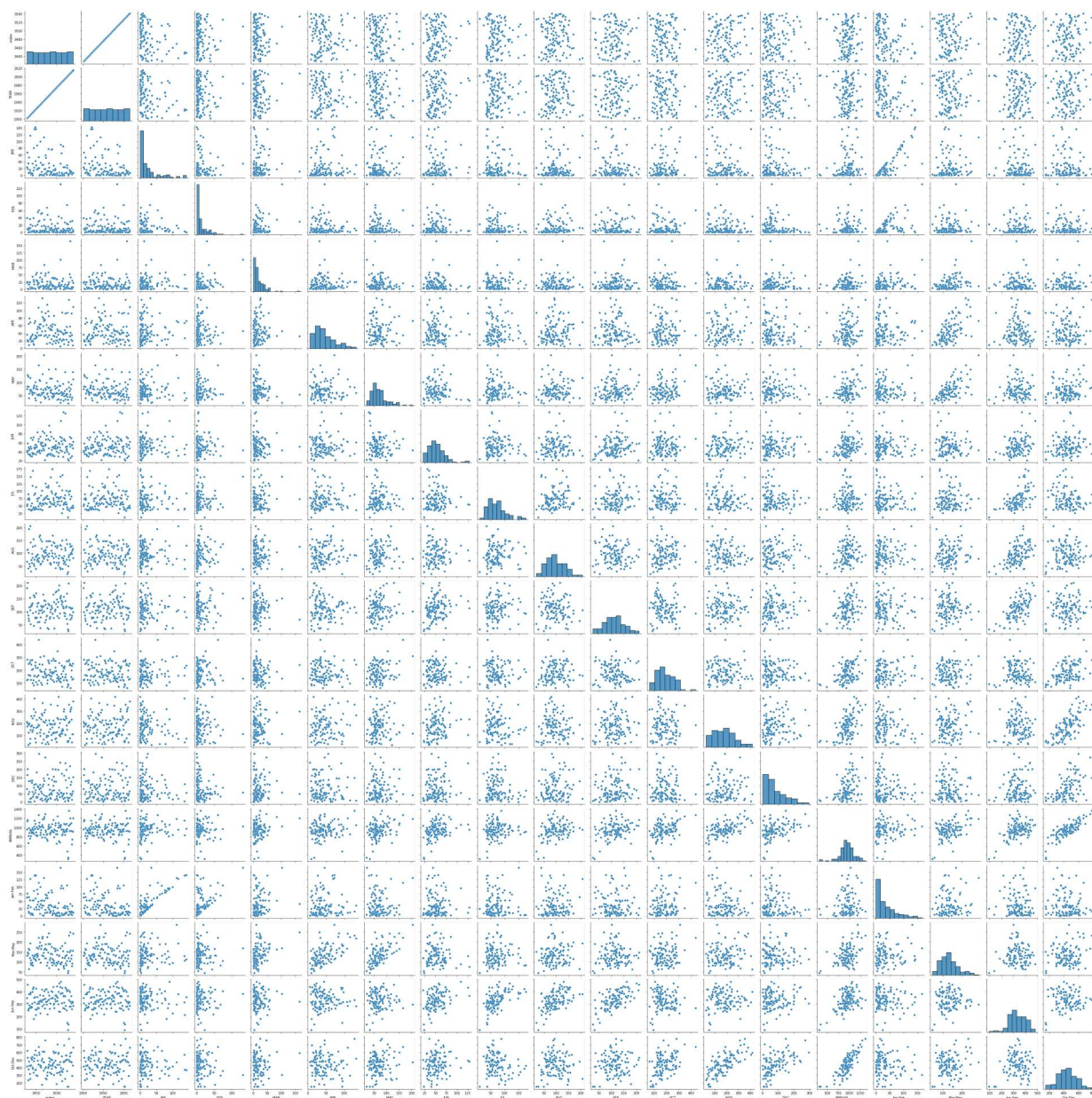
In [17]: `sns.displot(data["Jun-Sep"])`

Out[17]: `<seaborn.axisgrid.FacetGrid at 0x15b6fac8820>`

In [18]: `sns.heatmap(df1.corr())`

Out[18]: `<AxesSubplot:>`