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

LAKSHADWEEP

In [3]:
```
df=data.iloc[4002:4115]
df
```

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	
4002	4002	LAKSHADWEEP	1901	22.6	86.4	114.8	263.8	37.3	459.0	0.0	0.0	46.7	1
4003	4003	LAKSHADWEEP	1902	99.3	9.6	32.6	40.4	179.1	374.2	413.3	170.0	214.3	3
4004	4004	LAKSHADWEEP	1903	63.5	95.0	0.0	29.5	144.1	212.4	261.8	202.0	292.1	
4005	4005	LAKSHADWEEP	1904	0.0	0.0	13.5	13.2	143.3	261.3	256.0	38.9	219.9	1
4006	4006	LAKSHADWEEP	1905	62.4	0.0	0.0	0.0	166.7	400.7	68.7	377.5	107.5	2
...	...	...	...	...	...	...	...	...	...	...	...	...	
4110	4110	LAKSHADWEEP	2010	18.8	0.0	1.2	35.6	79.0	318.9	336.7	335.1	161.5	1
4111	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2	1
4112	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8	1
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	1

113 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	
4002	4002	LAKSHADWEEP	1901	22.6	86.4	114.8	263.8	37.3	459.0	0.0	0.0	46.7	1
4003	4003	LAKSHADWEEP	1902	99.3	9.6	32.6	40.4	179.1	374.2	413.3	170.0	214.3	3
4004	4004	LAKSHADWEEP	1903	63.5	95.0	0.0	29.5	144.1	212.4	261.8	202.0	292.1	
4005	4005	LAKSHADWEEP	1904	0.0	0.0	13.5	13.2	143.3	261.3	256.0	38.9	219.9	1
4006	4006	LAKSHADWEEP	1905	62.4	0.0	0.0	0.0	166.7	400.7	68.7	377.5	107.5	2

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

Out[5]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	O
4110	4110	LAKSHADWEEP	2010	18.8	0.0	1.2	35.6	79.0	318.9	336.7	335.1	161.5	15
4111	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2	11
4112	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8	14
4113	4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.0	7
4114	4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.2	16

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113 entries, 4002 to 4114
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   index        113 non-null    int64
 1   SUBDIVISION  113 non-null    object
 2   YEAR         113 non-null    int64
 3   JAN          111 non-null    float64
 4   FEB          112 non-null    float64
 5   MAR          111 non-null    float64
 6   APR          111 non-null    float64
 7   MAY          111 non-null    float64
 8   JUN          111 non-null    float64
 9   JUL          110 non-null    float64
 10  AUG          111 non-null    float64
 11  SEP          110 non-null    float64
 12  OCT          110 non-null    float64
 13  NOV          107 non-null    float64
 14  DEC          109 non-null    float64
 15  ANNUAL       102 non-null    float64
 16  Jan-Feb      110 non-null    float64
 17  Mar-May      109 non-null    float64
 18  Jun-Sep      109 non-null    float64
 19  Oct-Dec      107 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 17.8+ KB
```

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

Out[7]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	
4002	4002	LAKSHADWEEP	1901	22.6	86.4	114.8	263.8	37.3	459.0	0.0	0.0	46.7	1
4003	4003	LAKSHADWEEP	1902	99.3	9.6	32.6	40.4	179.1	374.2	413.3	170.0	214.3	3
4004	4004	LAKSHADWEEP	1903	63.5	95.0	0.0	29.5	144.1	212.4	261.8	202.0	292.1	
4005	4005	LAKSHADWEEP	1904	0.0	0.0	13.5	13.2	143.3	261.3	256.0	38.9	219.9	1
4006	4006	LAKSHADWEEP	1905	62.4	0.0	0.0	0.0	166.7	400.7	68.7	377.5	107.5	2
...	...	...	...	...	...	...	...	...	...	...	...	...	
4110	4110	LAKSHADWEEP	2010	18.8	0.0	1.2	35.6	79.0	318.9	336.7	335.1	161.5	1
4111	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2	1
4112	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8	1
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	1

113 rows × 20 columns

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

Out[8]:

|	index	YEAR	JAN	FEB	MAR	APR	MAY	
count	113.00000	113.000000	113.000000	113.000000	113.000000	113.000000	113.000000	113.0
mean	4058.00000	1957.849558	27.231858	15.830088	14.191150	43.992920	161.265487	322.1
std	32.76431	32.987882	38.075726	24.768539	21.290288	50.790293	110.891319	110.5
min	4002.00000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
25%	4030.00000	1930.000000	4.000000	0.400000	0.200000	12.300000	81.300000	251.2
50%	4058.00000	1958.000000	12.400000	4.100000	5.200000	32.000000	146.000000	326.1
75%	4086.00000	1986.000000	37.600000	18.500000	22.900000	57.600000	208.600000	380.5
max	4114.00000	2014.000000	262.800000	114.900000	120.700000	315.400000	660.800000	604.3

In [9]: `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'],`
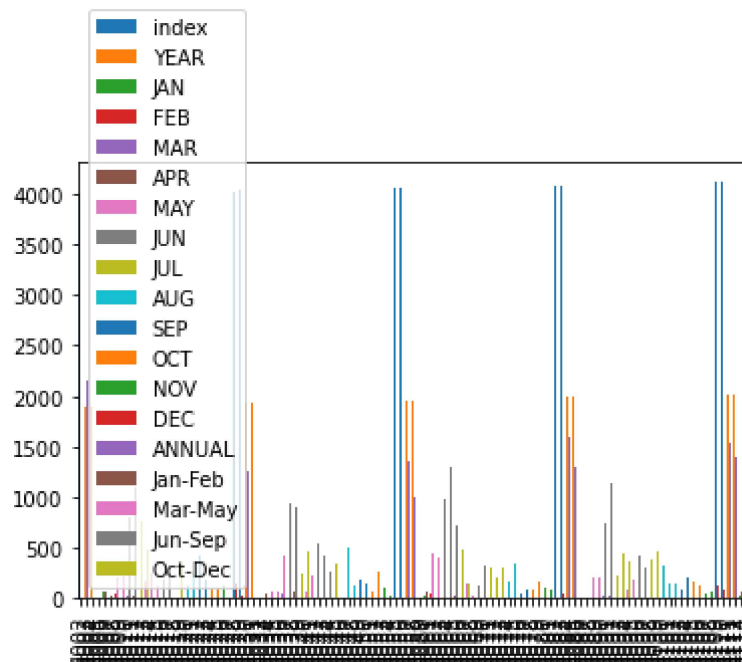`       dtype='object')`

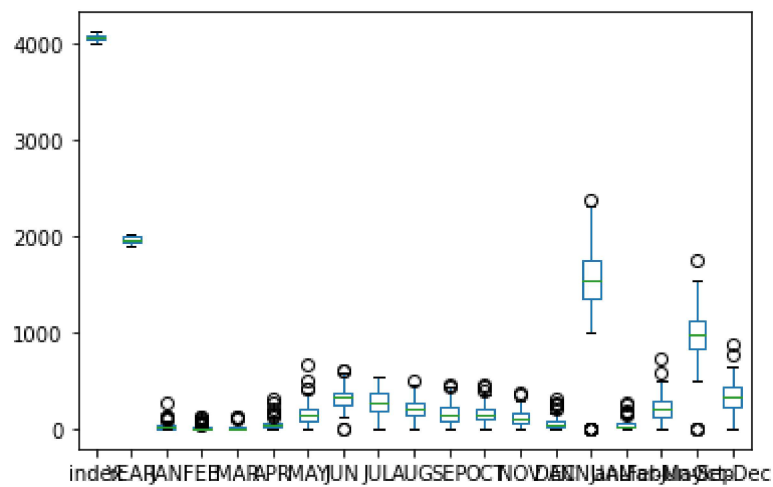# Data Visulaization

In [10]: `df1.plot.line()`

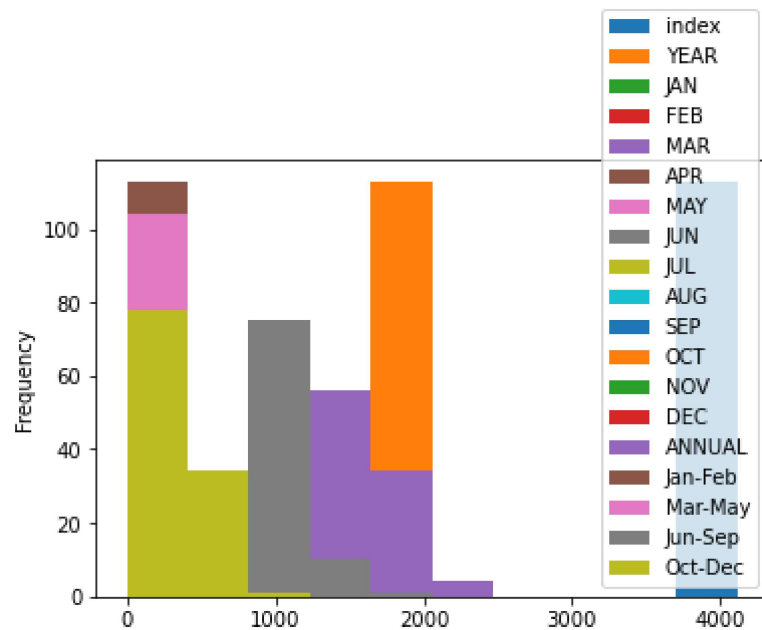Out[10]: `<AxesSubplot:>`
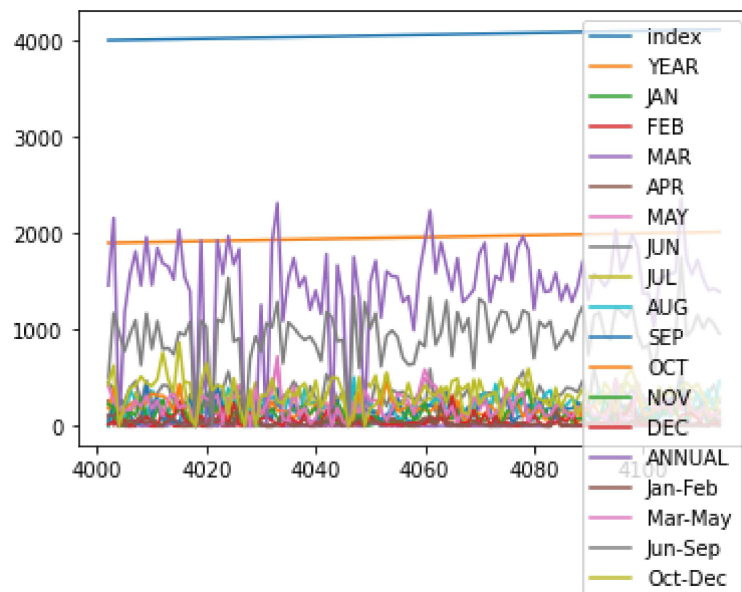
In [11]: `df1.plot.bar()`

Out[11]: `<AxesSubplot:>`

In [12]: `df1.plot.box()`

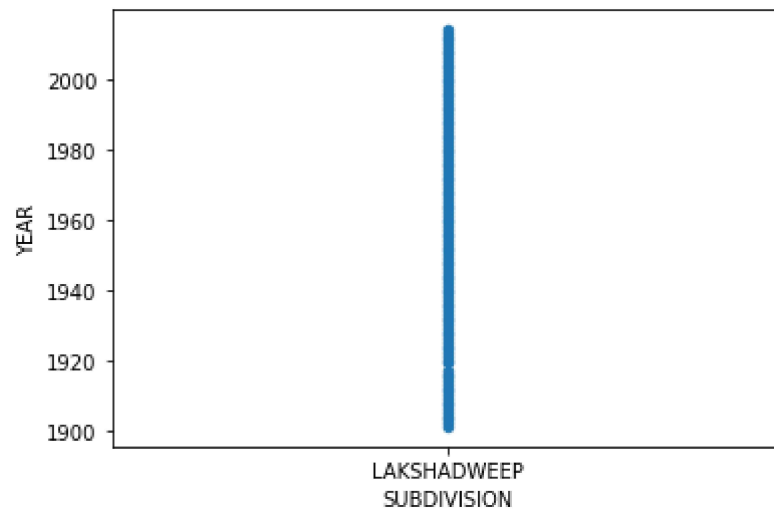Out[12]: `<AxesSubplot:>`

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

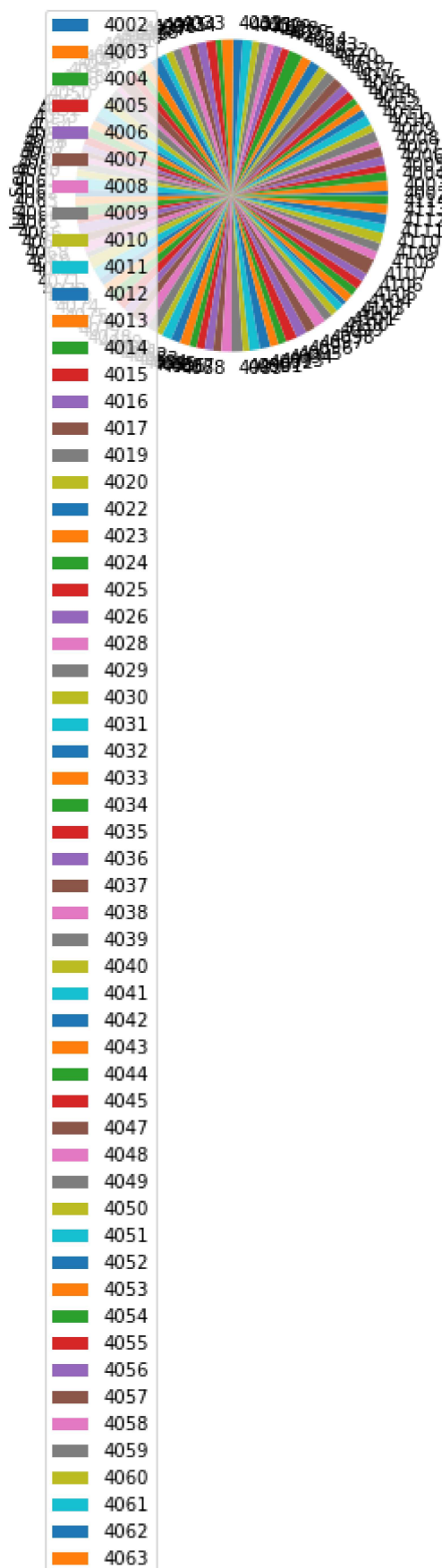Out[13]: `<AxesSubplot:ylabel='Frequency'>`

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

Out[14]: `<AxesSubplot:xlabel='SUBDIVISION', ylabel='YEAR'>`
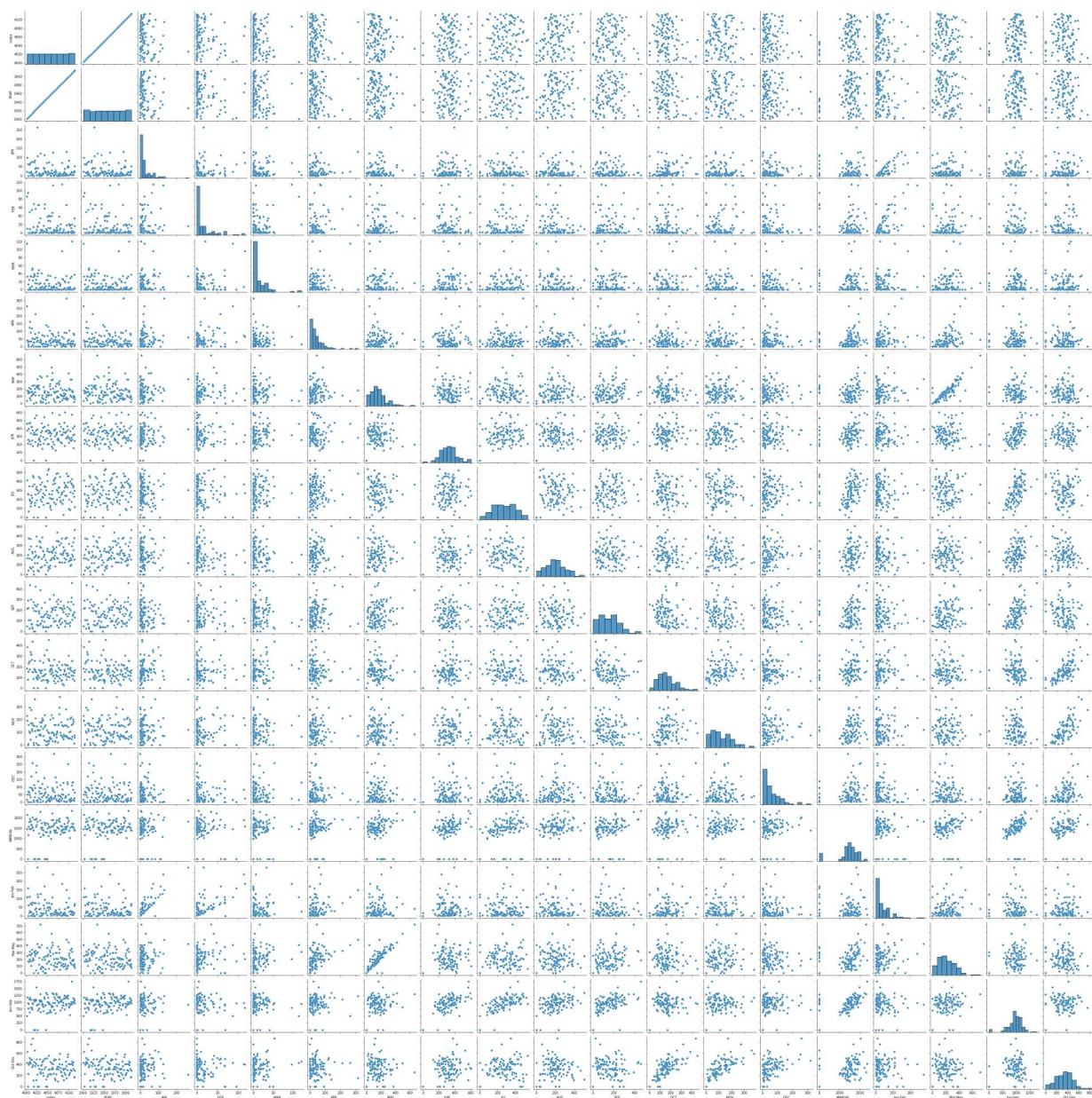
In [15]:
```python
df2=df1[[ 'Jun-Sep']]
df2.plot.pie(subplots=True)
```

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

	4064
	4065
	4066
	4067
	4068
	4069
	4070
	4071
	4072
	4073
	4074
	4075
	4076
	4077
	4078
	4079
	4080
	4081
	4082
	4083
	4084
	4085
	4086
	4087
	4088
	4089
	4090
	4091
	4092
	4093
	4094
	4095
	4096
	4097
	4098
	4099
	4100
	4101
	4102
	4103
	4104
	4105
	4106
	4107
	4108
	4109
	4110
	4111
	4112
	4113
	4114

In [16]: sns.pairplot(df1)

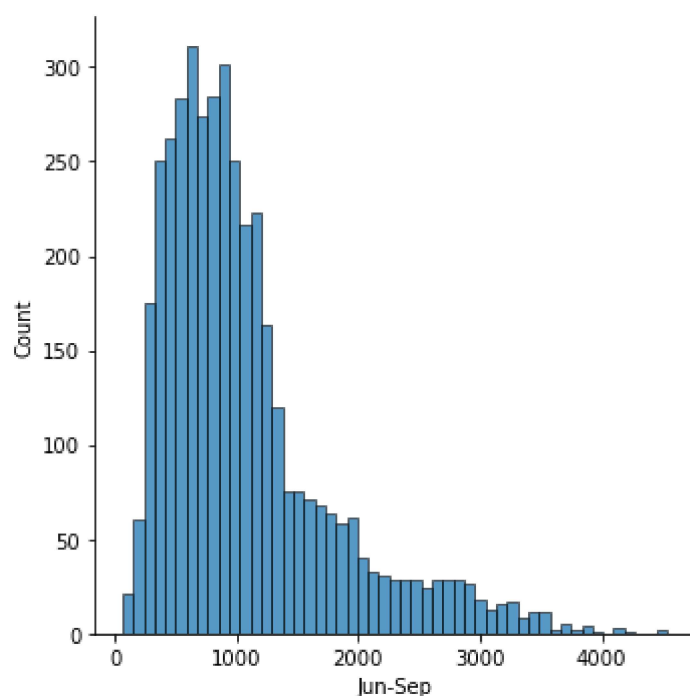Out[16]: <seaborn.axisgrid.PairGrid at 0x25cb4e84340>

In [17]: `sns.displot(data["Jun-Sep"])`

Out[17]: `<seaborn.axisgrid.FacetGrid at 0x25cb49528e0>`

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

Out[18]: `<AxesSubplot:>`