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

ARUNACHAL PRADESH

In [3]:
```
df=data.iloc[110:207]
df
```

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	
110	110	ARUNACHAL PRADESH	1916	48.1	69.8	71.1	316.1	424.6	1124.9	NaN	629.7	333.9	
111	111	ARUNACHAL PRADESH	1917	21.4	164.5	NaN	269.6	107.9	823.8	909.1	628.4	411.5	1
112	112	ARUNACHAL PRADESH	1918	10.4	11.0	191.2	144.6	861.1	1609.9	1303.0	692.6	515.8	1
113	113	ARUNACHAL PRADESH	1919	34.5	67.8	28.5	256.9	420.6	973.6	999.0	286.7	628.7	9
114	114	ARUNACHAL PRADESH	1920	14.0	196.3	605.6	364.7	173.6	840.6	535.4	896.5	376.7	1
...	...	...	...	...	...	...	...	...	...	...	...	...	
202	202	ARUNACHAL PRADESH	2011	40.0	51.3	174.5	240.8	219.6	288.4	531.4	277.6	286.7	
203	203	ARUNACHAL PRADESH	2012	57.8	35.8	134.2	403.4	187.4	645.8	638.9	316.0	724.9	2
204	204	ARUNACHAL PRADESH	2013	18.5	40.5	115.1	175.1	335.8	290.0	329.6	230.2	316.1	1
205	205	ARUNACHAL PRADESH	2014	19.0	101.9	80.3	86.7	299.0	415.8	392.4	599.6	343.0	
206	206	ARUNACHAL PRADESH	2015	30.8	47.5	97.5	287.1	238.9	637.9	329.3	595.5	374.2	

97 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	
110	110	ARUNACHAL PRADESH	1916	48.1	69.8	71.1	316.1	424.6	1124.9	NaN	629.7	333.9	
111	111	ARUNACHAL PRADESH	1917	21.4	164.5	NaN	269.6	107.9	823.8	909.1	628.4	411.5	1
112	112	ARUNACHAL PRADESH	1918	10.4	11.0	191.2	144.6	861.1	1609.9	1303.0	692.6	515.8	1
113	113	ARUNACHAL PRADESH	1919	34.5	67.8	28.5	256.9	420.6	973.6	999.0	286.7	628.7	9
114	114	ARUNACHAL PRADESH	1920	14.0	196.3	605.6	364.7	173.6	840.6	535.4	896.5	376.7	1

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

Out[5]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OC
202	202	ARUNACHAL PRADESH	2011	40.0	51.3	174.5	240.8	219.6	288.4	531.4	277.6	286.7	51
203	203	ARUNACHAL PRADESH	2012	57.8	35.8	134.2	403.4	187.4	645.8	638.9	316.0	724.9	248
204	204	ARUNACHAL PRADESH	2013	18.5	40.5	115.1	175.1	335.8	290.0	329.6	230.2	316.1	164
205	205	ARUNACHAL PRADESH	2014	19.0	101.9	80.3	86.7	299.0	415.8	392.4	599.6	343.0	35
206	206	ARUNACHAL PRADESH	2015	30.8	47.5	97.5	287.1	238.9	637.9	329.3	595.5	374.2	65

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 97 entries, 110 to 206
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   index        97 non-null     int64  
 1   SUBDIVISION  97 non-null     object 
 2   YEAR         97 non-null     int64  
 3   JAN          96 non-null     float64
 4   FEB          96 non-null     float64
 5   MAR          95 non-null     float64
 6   APR          97 non-null     float64
 7   MAY          97 non-null     float64
 8   JUN          96 non-null     float64
 9   JUL          96 non-null     float64
 10  AUG          97 non-null     float64
 11  SEP          97 non-null     float64
 12  OCT          95 non-null     float64
 13  NOV          95 non-null     float64
 14  DEC          95 non-null     float64
 15  ANNUAL       91 non-null     float64
 16  Jan-Feb      96 non-null     float64
 17  Mar-May      95 non-null     float64
 18  Jun-Sep      95 non-null     float64
 19  Oct-Dec      94 non-null     float64
dtypes: float64(17), int64(2), object(1)
memory usage: 15.3+ KB
```

In [19]:
```
#filling null values
df1=df.fillna(0)
df1
```

Out[19]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	
110	110	ARUNACHAL PRADESH	1916	48.1	69.8	71.1	316.1	424.6	1124.9	0.0	629.7	333.9	
111	111	ARUNACHAL PRADESH	1917	21.4	164.5	0.0	269.6	107.9	823.8	909.1	628.4	411.5	1
112	112	ARUNACHAL PRADESH	1918	10.4	11.0	191.2	144.6	861.1	1609.9	1303.0	692.6	515.8	1
113	113	ARUNACHAL PRADESH	1919	34.5	67.8	28.5	256.9	420.6	973.6	999.0	286.7	628.7	9
114	114	ARUNACHAL PRADESH	1920	14.0	196.3	605.6	364.7	173.6	840.6	535.4	896.5	376.7	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...
202	202	ARUNACHAL PRADESH	2011	40.0	51.3	174.5	240.8	219.6	288.4	531.4	277.6	286.7	
203	203	ARUNACHAL PRADESH	2012	57.8	35.8	134.2	403.4	187.4	645.8	638.9	316.0	724.9	2
204	204	ARUNACHAL PRADESH	2013	18.5	40.5	115.1	175.1	335.8	290.0	329.6	230.2	316.1	1
205	205	ARUNACHAL PRADESH	2014	19.0	101.9	80.3	86.7	299.0	415.8	392.4	599.6	343.0	
206	206	ARUNACHAL PRADESH	2015	30.8	47.5	97.5	287.1	238.9	637.9	329.3	595.5	374.2	

97 rows × 20 columns

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

Out[8]:

	index	YEAR	JAN	FEB	MAR	APR	MAY
count	4116.000000	4116.000000	4116.000000	4116.000000	4116.000000	4116.000000	4116.000000
mean	2057.500000	1958.218659	18.938897	21.789431	27.319315	43.085520	85.682920
std	1188.331183	33.140898	33.574242	35.901220	46.936787	67.811512	123.211711
min	0.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1028.750000	1930.000000	0.600000	0.600000	1.000000	3.000000	8.600000
50%	2057.500000	1958.000000	6.000000	6.700000	7.800000	15.600000	36.400000
75%	3086.250000	1987.000000	22.125000	26.800000	31.225000	49.825000	96.825000
max	4115.000000	2015.000000	583.700000	403.500000	605.600000	595.100000	1168.600000

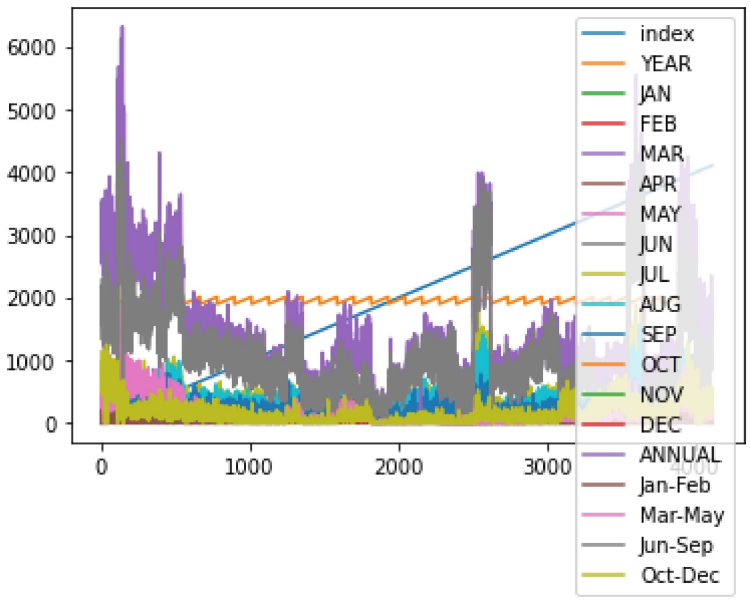
In [9]:
```
df1.columns
```

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

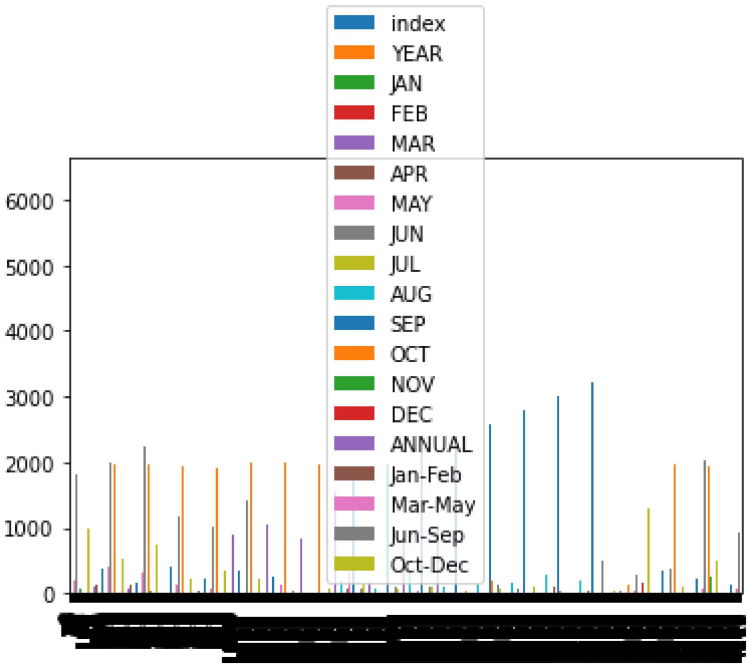
In [10]:
```
df1.plot.line()
```

Out[10]:
```
<AxesSubplot:>
```



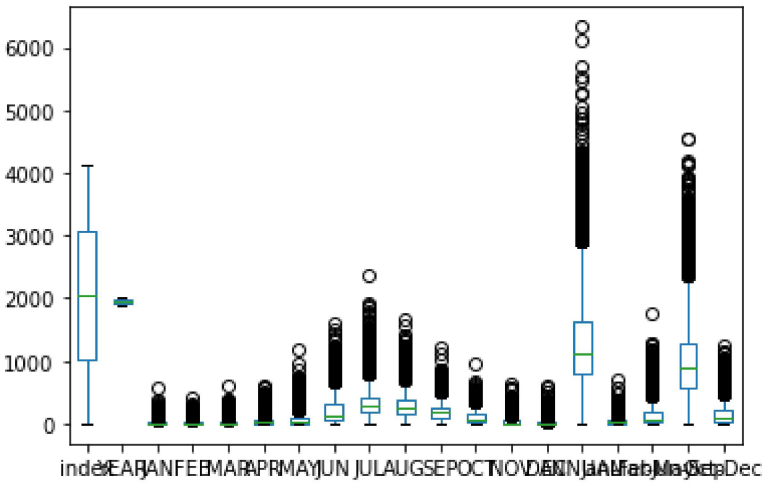
In [11]:
```
df1.plot.bar()
```

Out[11]: <AxesSubplot:>



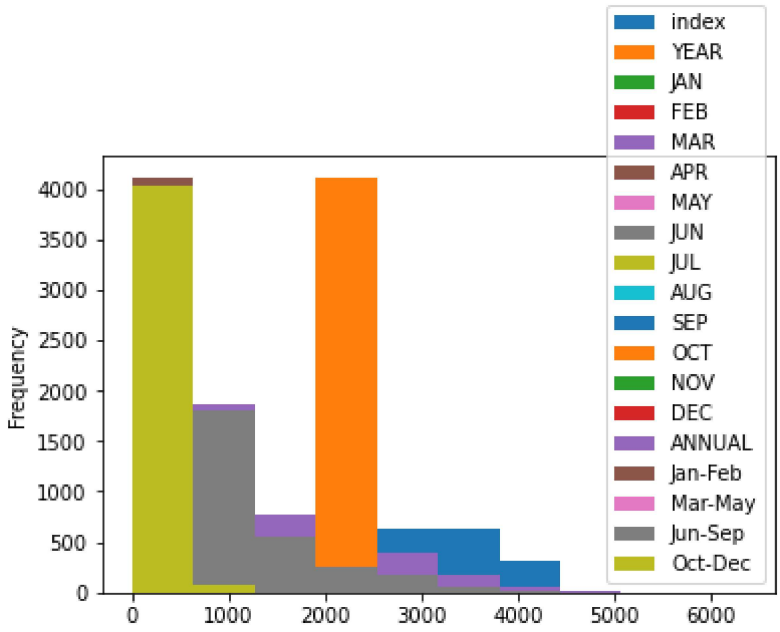
In [12]:
```
df1.plot.box()
```

Out[12]: <AxesSubplot:>



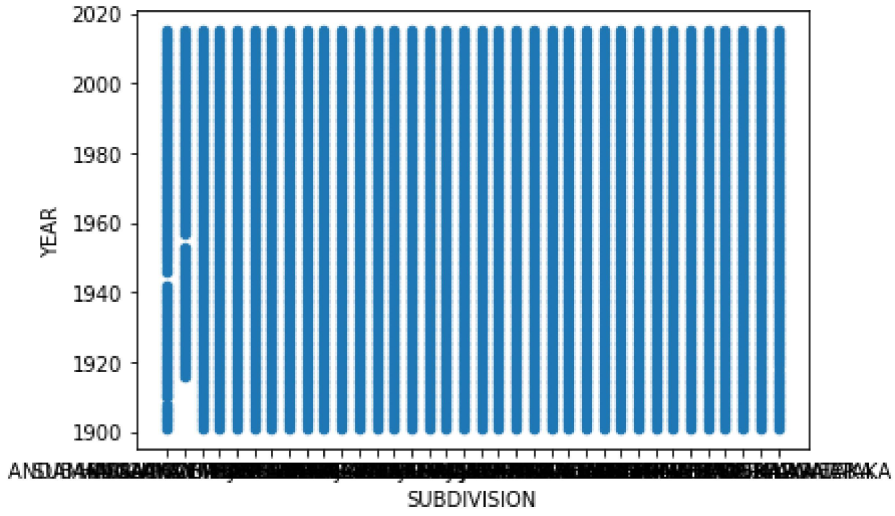
In [13]:
```
df1.plot.hist()
```

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

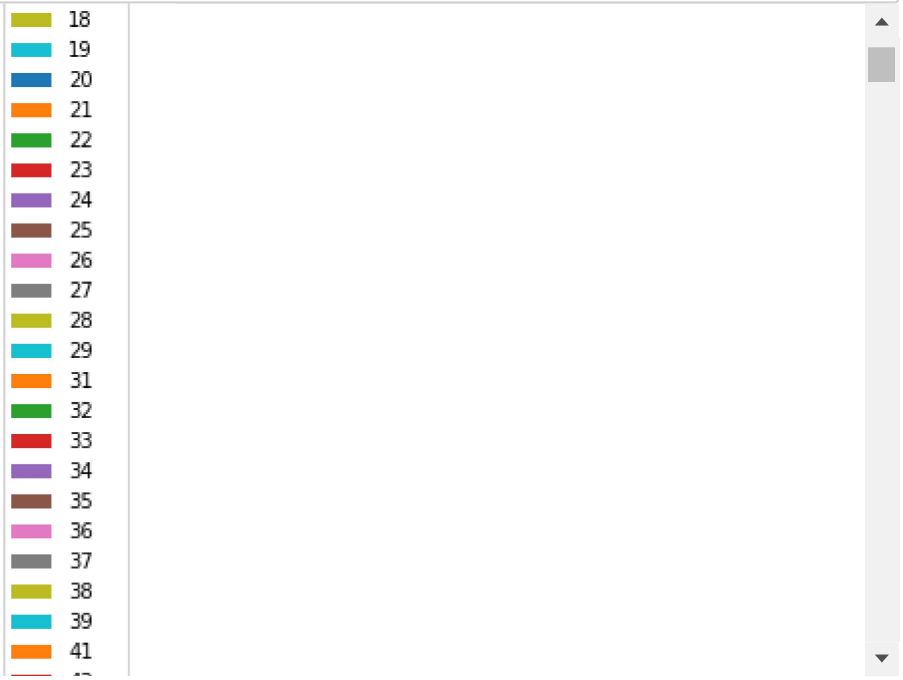


In [14]:
```
df1.plot.scatter(x="SUBDIVISION",y="YEAR")
```

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

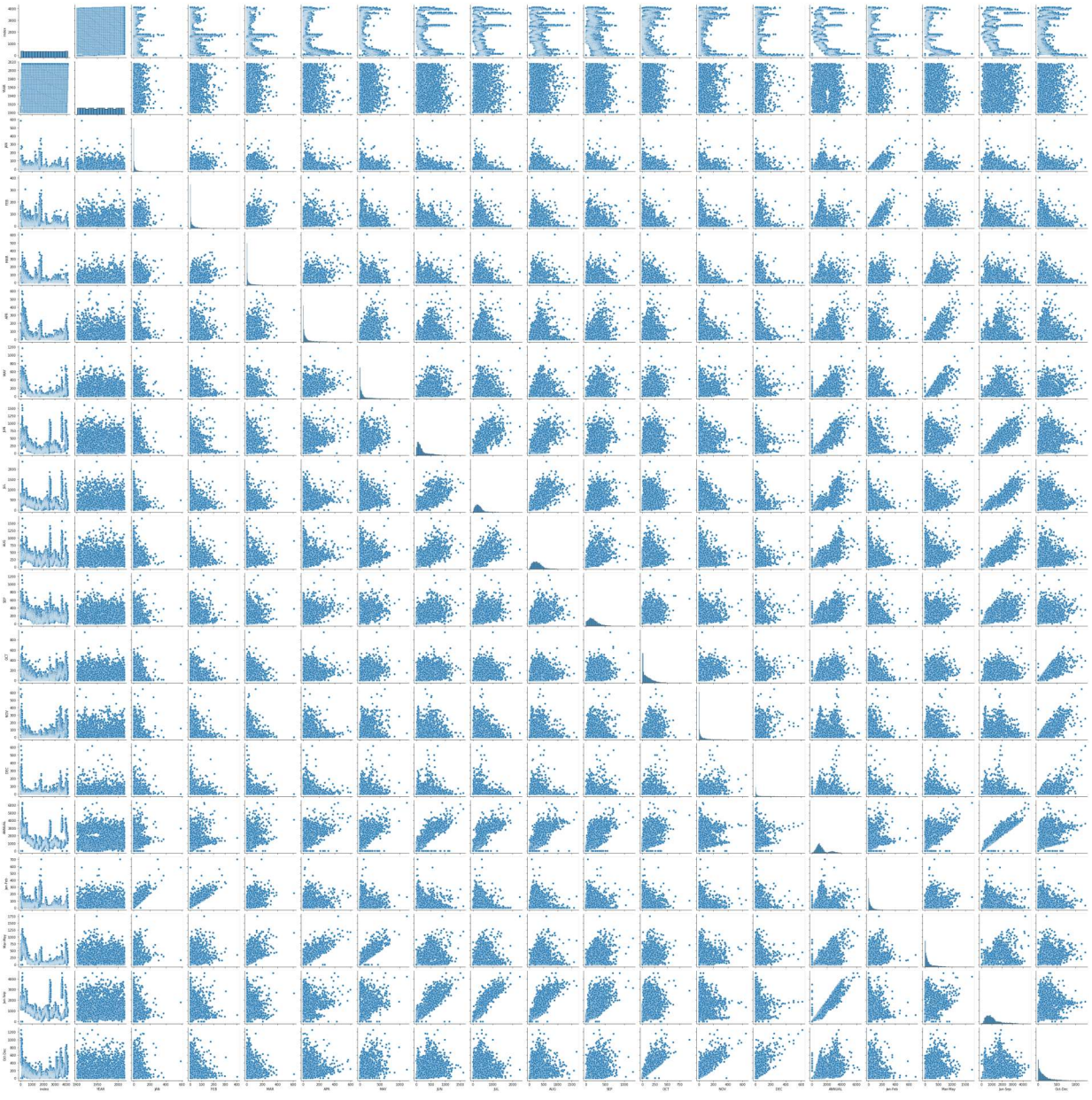


In [15]:
```
df2=df1[[ 'Jun-Sep']]
df2.plot.pie(subplots=True)
```



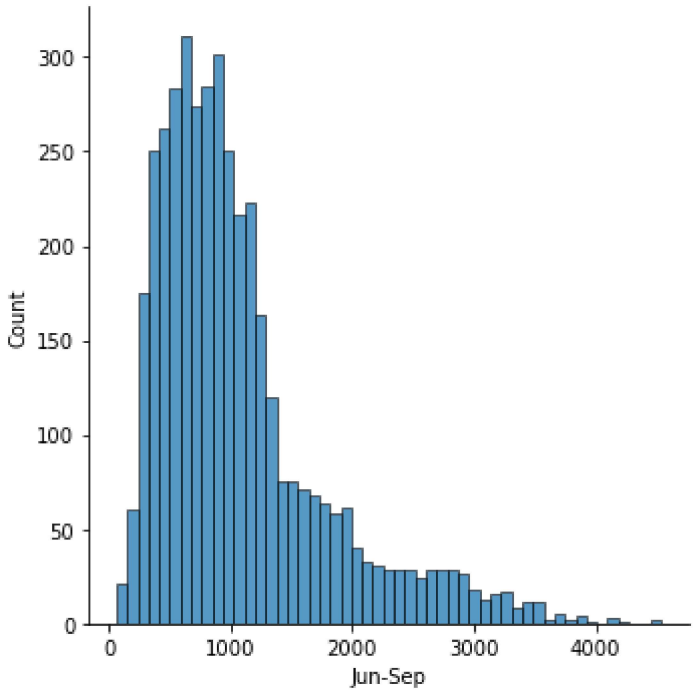
In [16]:
```
sns.pairplot(df1)
```

Out[16]: <seaborn.axisgrid.PairGrid at 0x1cc39915f70>



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

Out[17]:
```
<seaborn.axisgrid.FacetGrid at 0x1cc667dd370>
```



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

Out[18]: <AxesSubplot:>