

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

```
In [2]: 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



Assam and Megayala

In [3]:

df=data.iloc[207:321]
df

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	O
207	207	ASSAM & MEGHALAYA	1901	27.1	19.5	30.6	223.0	207.0	524.9	430.6	464.1	291.4	16
208	208	ASSAM & MEGHALAYA	1902	9.3	10.2	105.6	350.0	262.1	620.7	510.8	536.0	441.3	9
209	209	ASSAM & MEGHALAYA	1903	19.9	25.4	103.6	140.6	206.6	607.4	362.7	551.9	306.4	15
210	210	ASSAM & MEGHALAYA	1904	11.1	56.1	51.9	457.1	375.2	385.7	477.6	438.8	245.9	11
211	211	ASSAM & MEGHALAYA	1905	19.9	16.9	137.9	213.0	275.5	521.7	439.1	649.1	276.0	20
...
316	316	ASSAM & MEGHALAYA	2010	0.5	3.1	99.6	389.3	393.4	569.7	415.1	387.0	318.3	11
317	317	ASSAM & MEGHALAYA	2011	11.1	11.4	109.0	92.1	238.3	316.0	395.8	302.6	221.6	3
318	318	ASSAM & MEGHALAYA	2012	15.2	6.9	28.8	279.1	185.8	729.7	444.3	289.2	411.6	19
319	319	ASSAM & MEGHALAYA	2013	1.1	9.6	44.0	112.8	346.7	286.2	367.8	289.7	229.3	12
320	320	ASSAM & MEGHALAYA	2014	2.0	28.3	29.3	51.5	351.1	426.4	374.4	484.6	420.2	3

114 rows × 20 columns

In [4]:

df.head()

Out[4]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	O
207	207	ASSAM & MEGHALAYA	1901	27.1	19.5	30.6	223.0	207.0	524.9	430.6	464.1	291.4	16
208	208	ASSAM & MEGHALAYA	1902	9.3	10.2	105.6	350.0	262.1	620.7	510.8	536.0	441.3	9
209	209	ASSAM & MEGHALAYA	1903	19.9	25.4	103.6	140.6	206.6	607.4	362.7	551.9	306.4	15
210	210	ASSAM & MEGHALAYA	1904	11.1	56.1	51.9	457.1	375.2	385.7	477.6	438.8	245.9	11
211	211	ASSAM & MEGHALAYA	1905	19.9	16.9	137.9	213.0	275.5	521.7	439.1	649.1	276.0	20

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

Out[5]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
316	316	ASSAM & MEGHALAYA	2010	0.5	3.1	99.6	389.3	393.4	569.7	415.1	387.0	318.3	116
317	317	ASSAM & MEGHALAYA	2011	11.1	11.4	109.0	92.1	238.3	316.0	395.8	302.6	221.6	30
318	318	ASSAM & MEGHALAYA	2012	15.2	6.9	28.8	279.1	185.8	729.7	444.3	289.2	411.6	196
319	319	ASSAM & MEGHALAYA	2013	1.1	9.6	44.0	112.8	346.7	286.2	367.8	289.7	229.3	126
320	320	ASSAM & MEGHALAYA	2014	2.0	28.3	29.3	51.5	351.1	426.4	374.4	484.6	420.2	31

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 114 entries, 207 to 320
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
```

```
In [7]: df1=data.fillna(0)
df1
```

Out[7]:

	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

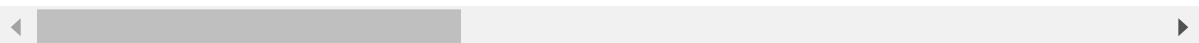
4116 rows × 20 columns



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

Out[8]:

	index	YEAR	JAN	FEB	MAR	APR	MA
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	123.211700
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

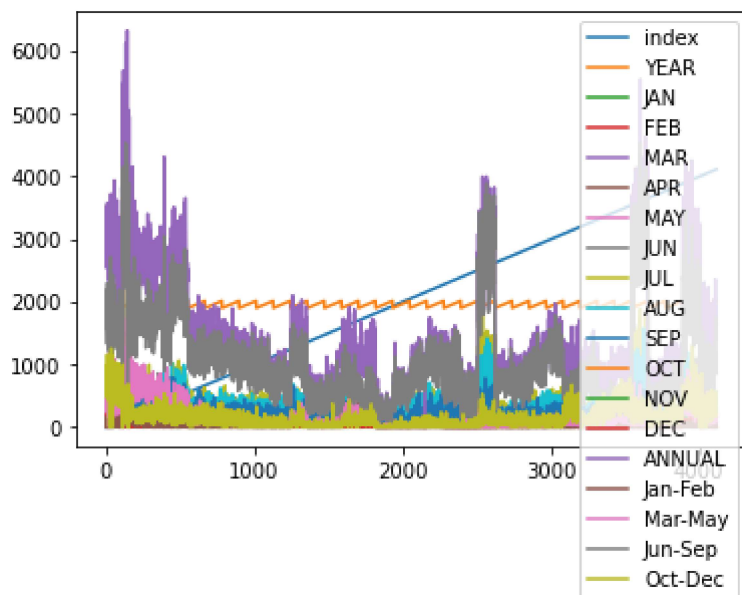


```
In [9]: df1.columns
```

```
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')
```

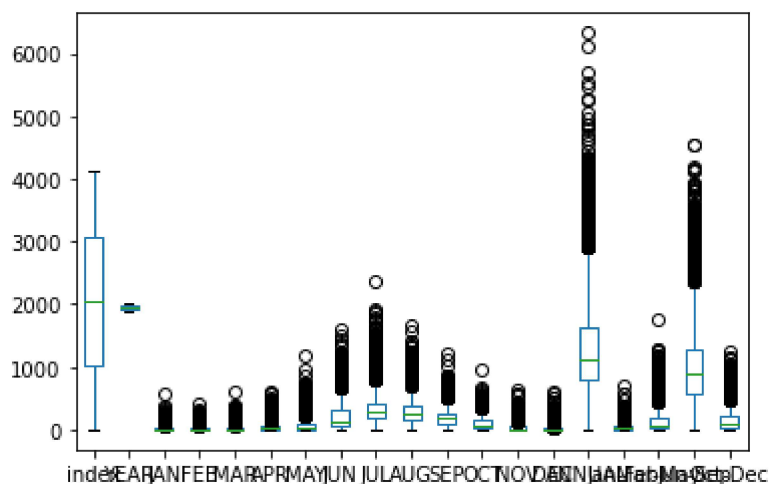
```
In [10]: df1.plot.line()
```

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



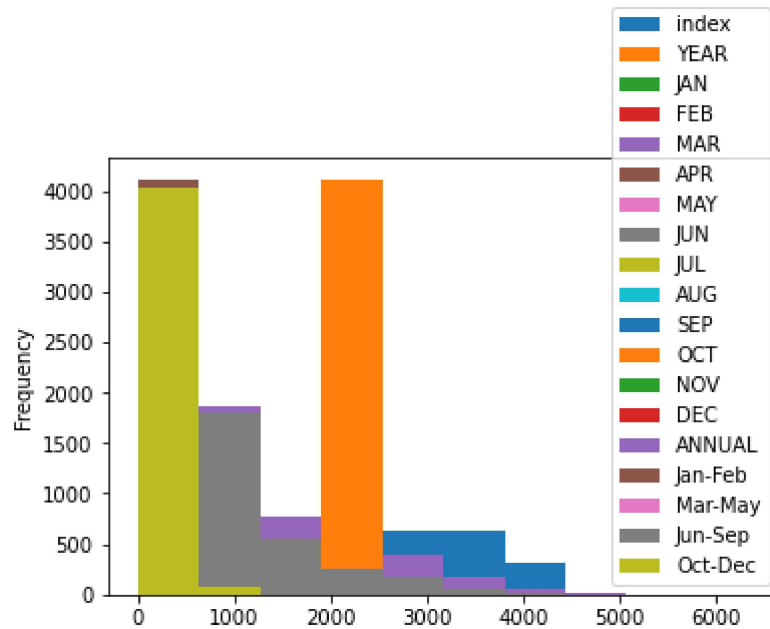
```
In [11]: df1.plot.box()
```

```
Out[11]: <AxesSubplot:>
```



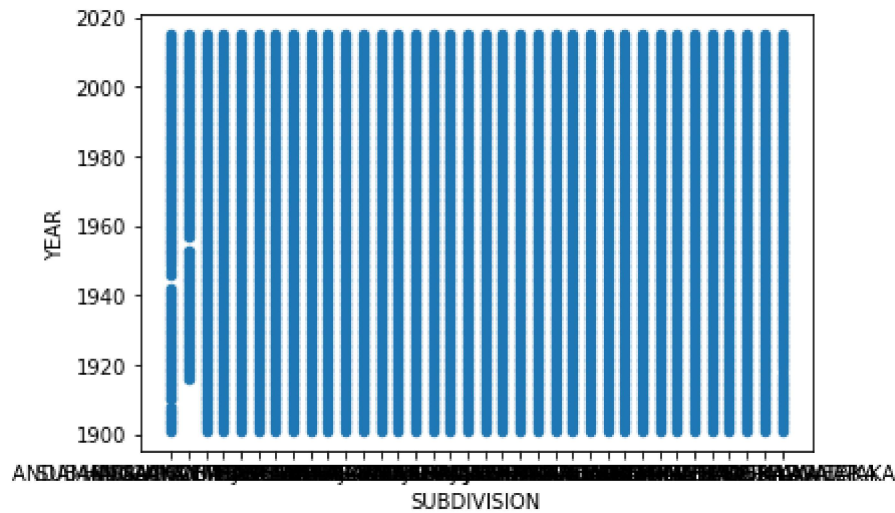
```
In [12]: df1.plot.hist()
```

```
Out[12]: <AxesSubplot:ylabel='Frequency'>
```



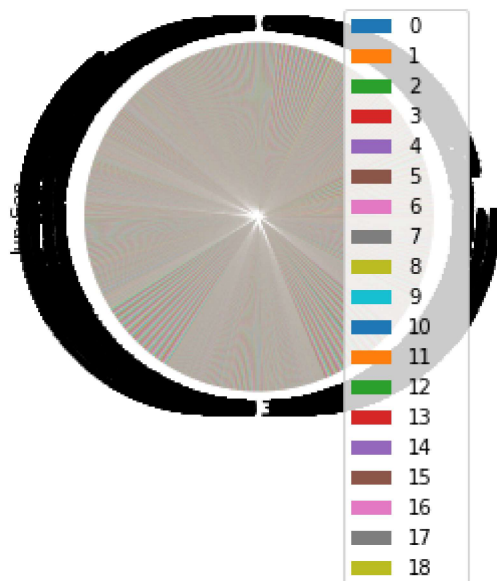
```
In [13]: df1.plot.scatter(x="SUBDIVISION",y="YEAR")
```

```
Out[13]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='YEAR'>
```



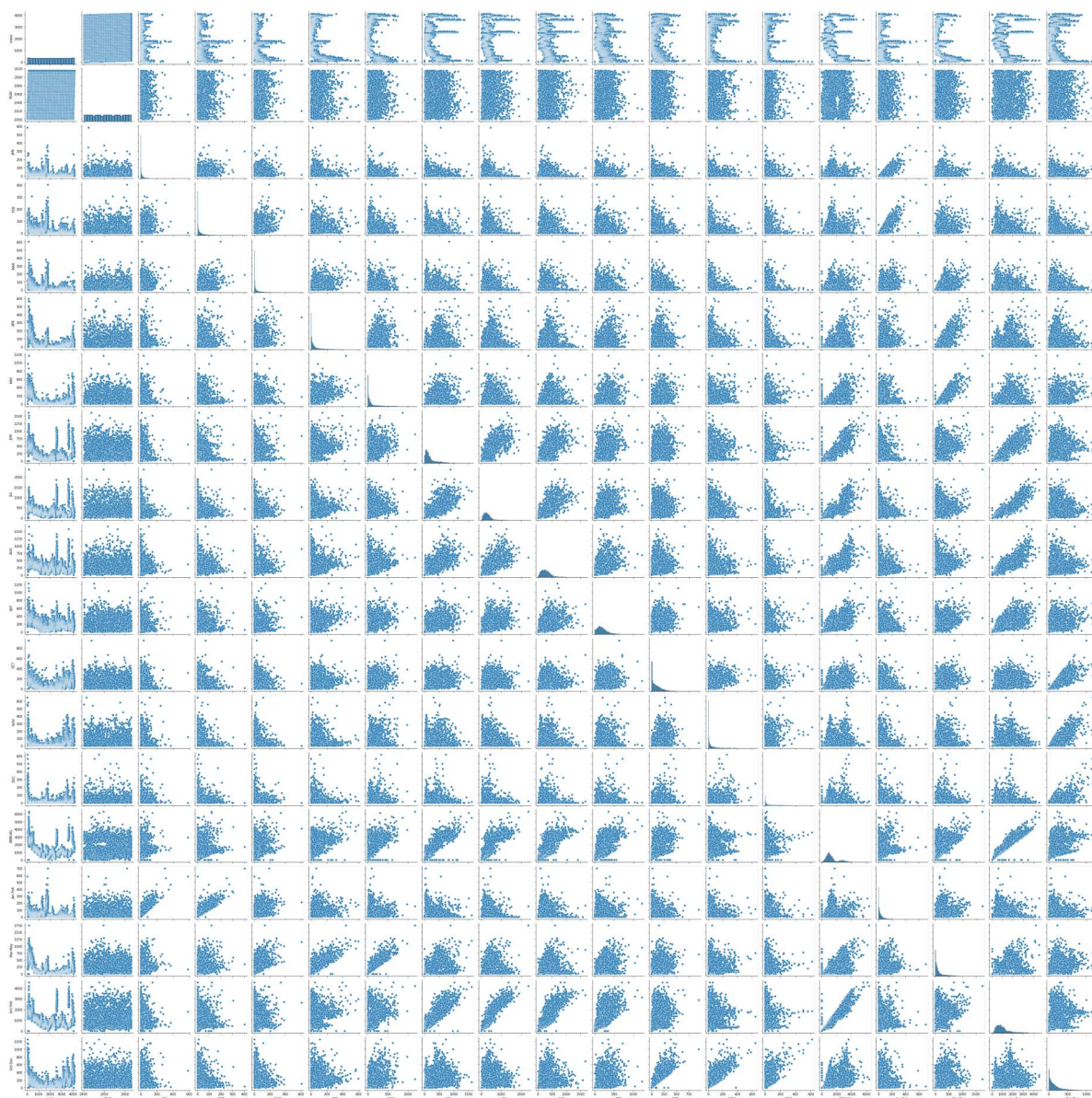
```
In [14]: df2=df1[['Jun-Sep']]  
df2.plot.pie(subplots=True)
```

```
Out[14]: array([<AxesSubplot:ylabel='Jun-Sep'>], dtype=object)
```



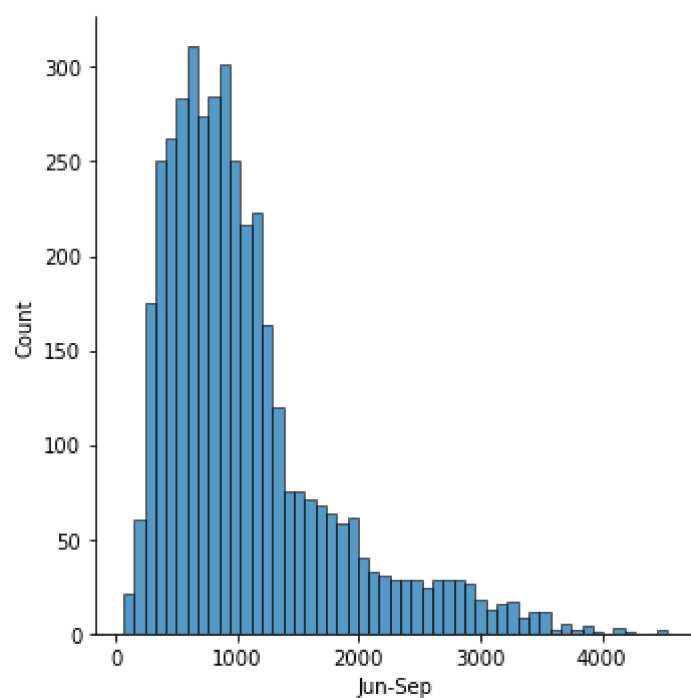

```
In [18]: sns.pairplot(df1)
```

```
Out[18]: <seaborn.axisgrid.PairGrid at 0x1fbf5d18ee0>
```



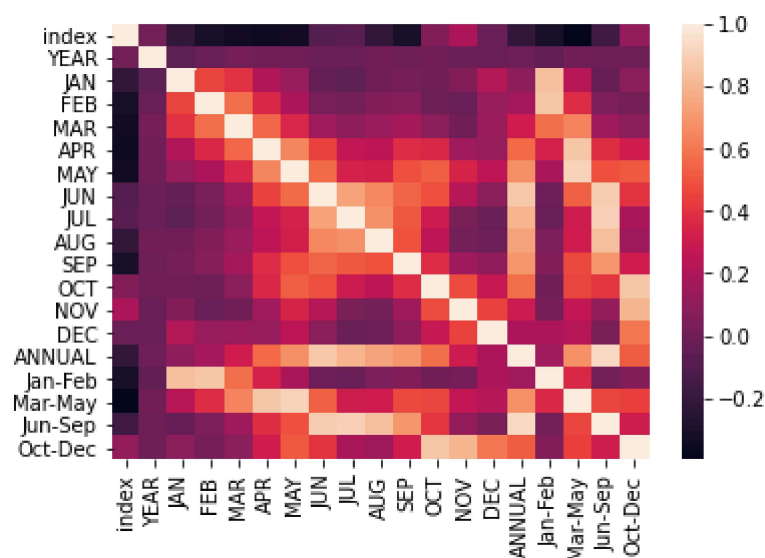

```
In [19]: sns.displot(data["Jun-Sep"])
```

```
Out[19]: <seaborn.axisgrid.FacetGrid at 0x1fba35172e0>
```



```
In [20]: sns.heatmap(df1.corr())
```

```
Out[20]: <AxesSubplot:>
```



```
In [ ]:
```