

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

COASTAL KARNATAKA

```
In [3]: df=data.iloc[3542:3658]
df
```

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	C
3542	3542	COASTAL KARNATAKA	1901	1.8	0.6	10.7	52.4	81.6	960.9	991.2	606.4	108.0	12
3543	3543	COASTAL KARNATAKA	1902	3.2	0.3	4.9	10.2	54.6	698.4	1401.6	454.2	708.4	18
3544	3544	COASTAL KARNATAKA	1903	0.7	0.0	0.0	4.1	202.8	536.5	1405.5	593.8	304.4	18
3545	3545	COASTAL KARNATAKA	1904	2.4	0.0	4.8	23.7	93.2	1108.2	1070.0	465.6	245.3	12
3546	3546	COASTAL KARNATAKA	1905	0.0	0.2	0.0	6.4	83.1	767.3	777.3	586.9	172.9	22
...
3653	3653	COASTAL KARNATAKA	2012	NaN	11.4	5.1	77.0	22.9	650.9	754.6	1027.6	382.0	11
3654	3654	COASTAL KARNATAKA	2013	2.4	19.6	19.0	28.5	100.4	1153.0	1515.3	680.2	379.1	26
3655	3655	COASTAL KARNATAKA	2014	0.0	0.3	1.9	40.5	181.9	507.0	1155.4	1121.0	379.3	22
3656	3656	COASTAL KARNATAKA	2015	1.4	1.0	32.3	72.2	150.3	735.3	930.9	575.2	260.3	20
3657	3657	NORTH INTERIOR KARNATAKA	1901	3.5	18.8	7.1	67.2	65.5	120.5	151.9	115.1	128.8	8

116 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	O
3542	3542	COASTAL KARNATAKA	1901	1.8	0.6	10.7	52.4	81.6	960.9	991.2	606.4	108.0	120
3543	3543	COASTAL KARNATAKA	1902	3.2	0.3	4.9	10.2	54.6	698.4	1401.6	454.2	708.4	180
3544	3544	COASTAL KARNATAKA	1903	0.7	0.0	0.0	4.1	202.8	536.5	1405.5	593.8	304.4	180
3545	3545	COASTAL KARNATAKA	1904	2.4	0.0	4.8	23.7	93.2	1108.2	1070.0	465.6	245.3	120
3546	3546	COASTAL KARNATAKA	1905	0.0	0.2	0.0	6.4	83.1	767.3	777.3	586.9	172.9	220

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

Out[5]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	C
3653	3653	COASTAL KARNATAKA	2012	NaN	11.4	5.1	77.0	22.9	650.9	754.6	1027.6	382.0	10
3654	3654	COASTAL KARNATAKA	2013	2.4	19.6	19.0	28.5	100.4	1153.0	1515.3	680.2	379.1	20
3655	3655	COASTAL KARNATAKA	2014	0.0	0.3	1.9	40.5	181.9	507.0	1155.4	1121.0	379.3	20
3656	3656	COASTAL KARNATAKA	2015	1.4	1.0	32.3	72.2	150.3	735.3	930.9	575.2	260.3	20
3657	3657	NORTH INTERIOR KARNATAKA	1901	3.5	18.8	7.1	67.2	65.5	120.5	151.9	115.1	128.8	8

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 116 entries, 3542 to 3657
Data columns (total 20 columns):
#   Column          Non-Null Count  Dtype
---  -
0   index           116 non-null    int64
1   SUBDIVISION     116 non-null    object
2   YEAR            116 non-null    int64
3   JAN             115 non-null    float64
4   FEB             116 non-null    float64
5   MAR             116 non-null    float64
6   APR             116 non-null    float64
7   MAY             116 non-null    float64
8   JUN             116 non-null    float64
9   JUL             116 non-null    float64
10  AUG             116 non-null    float64
11  SEP             116 non-null    float64
12  OCT             116 non-null    float64
13  NOV             116 non-null    float64
14  DEC             116 non-null    float64
15  ANNUAL          115 non-null    float64
16  Jan-Feb         115 non-null    float64
17  Mar-May         116 non-null    float64
18  Jun-Sep         116 non-null    float64
19  Oct-Dec         116 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 18.3+ 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	C
3542	3542	COASTAL KARNATAKA	1901	1.8	0.6	10.7	52.4	81.6	960.9	991.2	606.4	108.0	12
3543	3543	COASTAL KARNATAKA	1902	3.2	0.3	4.9	10.2	54.6	698.4	1401.6	454.2	708.4	18
3544	3544	COASTAL KARNATAKA	1903	0.7	0.0	0.0	4.1	202.8	536.5	1405.5	593.8	304.4	18
3545	3545	COASTAL KARNATAKA	1904	2.4	0.0	4.8	23.7	93.2	1108.2	1070.0	465.6	245.3	12
3546	3546	COASTAL KARNATAKA	1905	0.0	0.2	0.0	6.4	83.1	767.3	777.3	586.9	172.9	22
...
3653	3653	COASTAL KARNATAKA	2012	0.0	11.4	5.1	77.0	22.9	650.9	754.6	1027.6	382.0	17
3654	3654	COASTAL KARNATAKA	2013	2.4	19.6	19.0	28.5	100.4	1153.0	1515.3	680.2	379.1	26
3655	3655	COASTAL KARNATAKA	2014	0.0	0.3	1.9	40.5	181.9	507.0	1155.4	1121.0	379.3	22
3656	3656	COASTAL KARNATAKA	2015	1.4	1.0	32.3	72.2	150.3	735.3	930.9	575.2	260.3	20
3657	3657	NORTH INTERIOR KARNATAKA	1901	3.5	18.8	7.1	67.2	65.5	120.5	151.9	115.1	128.8	8

116 rows × 20 columns

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

Out[8]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	
count	116.000000	116.000000	116.000000	116.000000	116.000000	116.000000	116.000000	116.000000
mean	3599.500000	1957.508621	1.934483	1.667241	6.363793	31.22931	122.293966	835.000000
std	33.630343	33.615601	4.187956	4.974540	16.501264	24.26323	124.960178	203.000000
min	3542.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	8.400000	120.000000
25%	3570.750000	1928.750000	0.000000	0.000000	0.200000	11.37500	43.125000	699.000000
50%	3599.500000	1957.500000	0.100000	0.000000	1.550000	24.95000	80.150000	810.000000
75%	3628.250000	1986.250000	2.000000	0.600000	6.300000	45.40000	158.675000	976.000000
max	3657.000000	2015.000000	23.000000	29.800000	161.400000	110.10000	699.500000	1361.000000

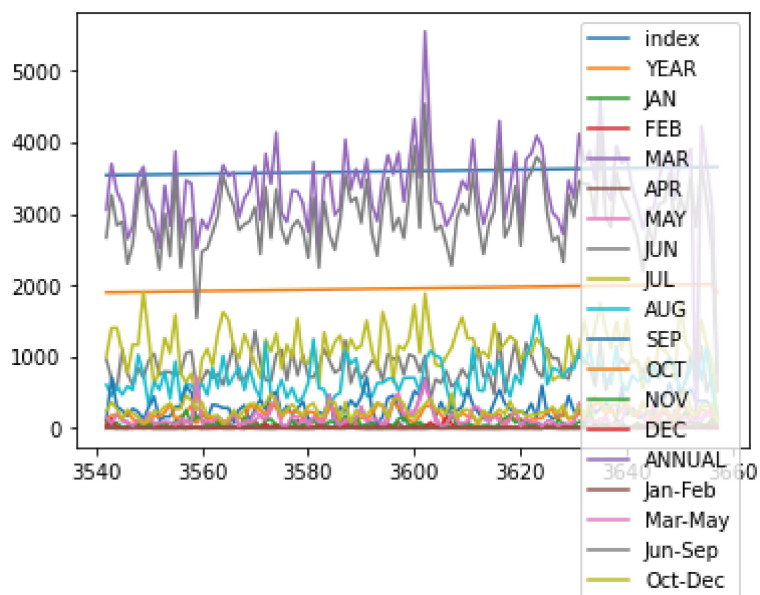
In [9]: `df1.columns`

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

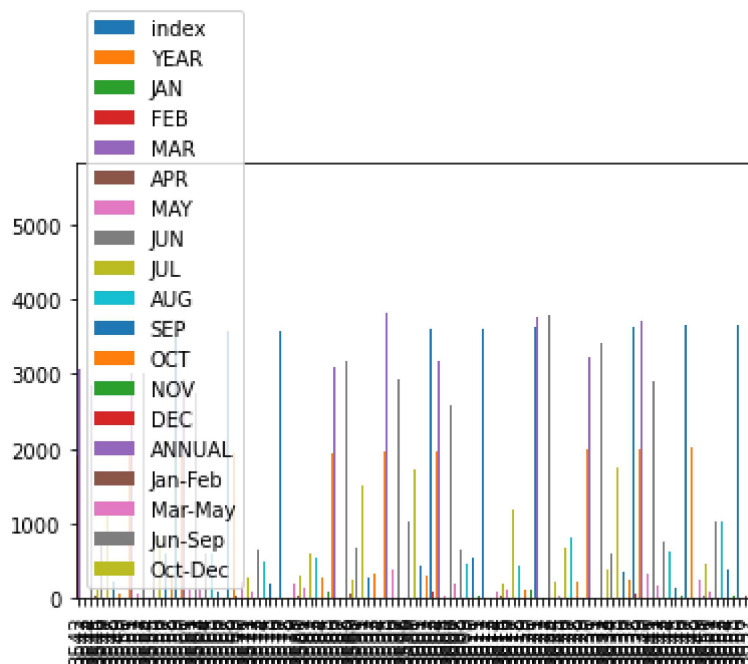
In [10]: `df1.plot.line()`

Out[10]: <AxesSubplot:>



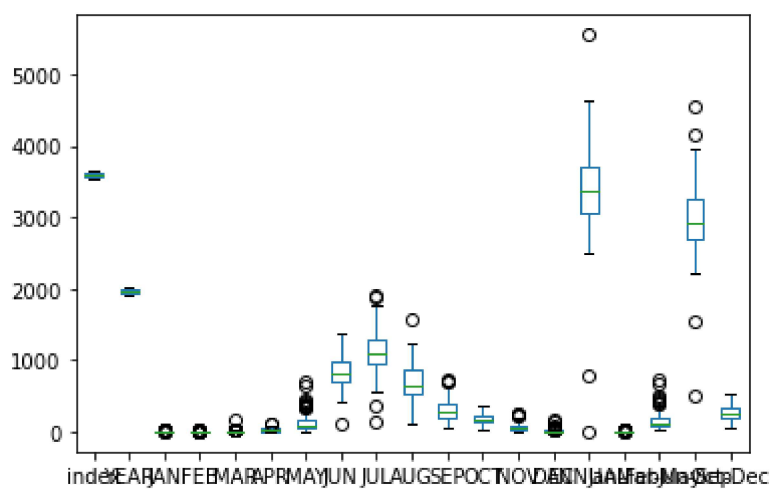
```
In [11]: df1.plot.bar()
```

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



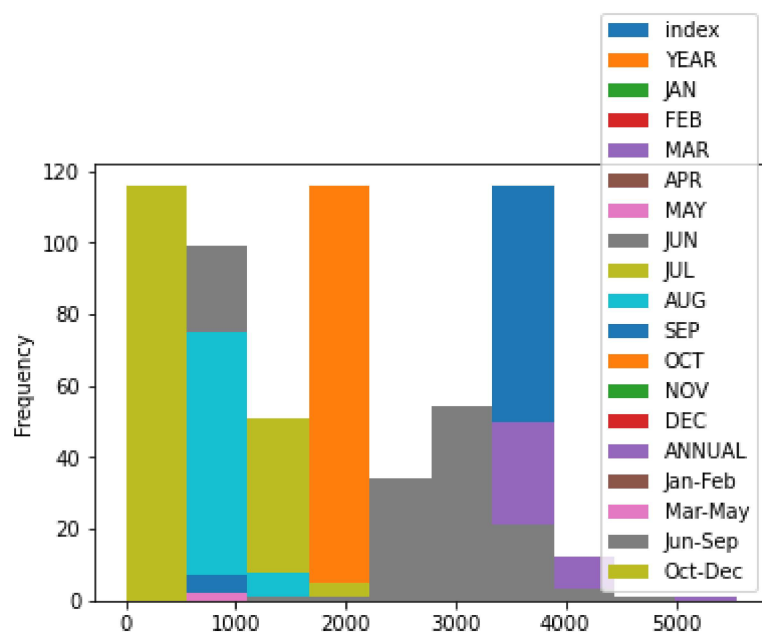
```
In [12]: df1.plot.box()
```

```
Out[12]: <AxesSubplot:>
```



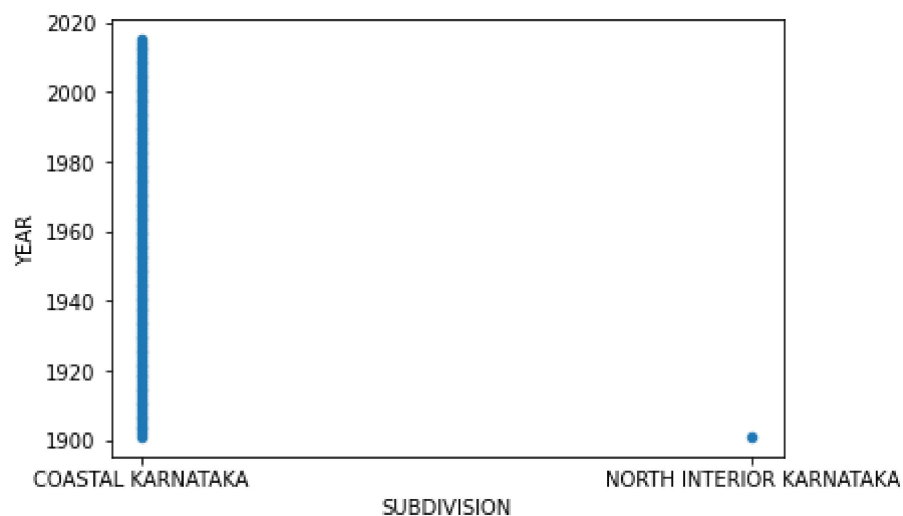
```
In [13]: df1.plot.hist()
```

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



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

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

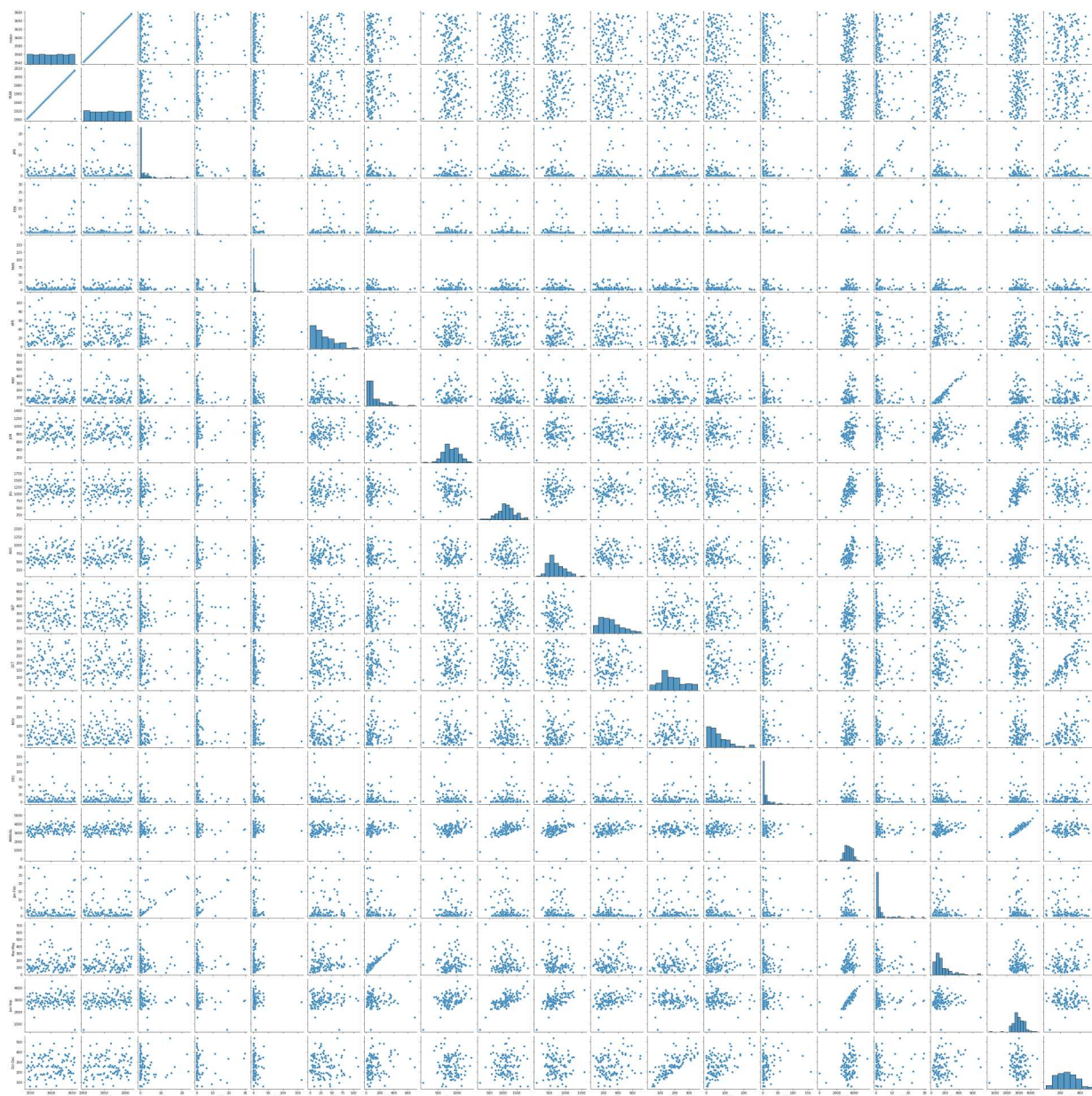



```
In [15]: df2=df1[[ 'Jun-Sep']]  
df2.plot.pie(subplots=True)
```

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

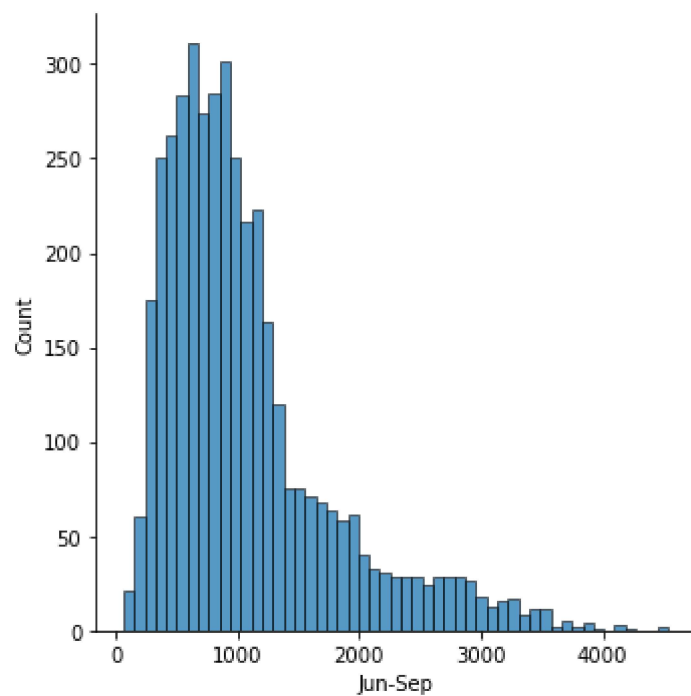
```
In [16]: sns.pairplot(df1)
```

```
Out[16]: <seaborn.axisgrid.PairGrid at 0x21c94c36730>
```



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

```
Out[17]: <seaborn.axisgrid.FacetGrid at 0x21ca34d38e0>
```



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

```
Out[18]: <AxesSubplot:>
```

