In [1]: import pandas as pd
data=pd.read\_csv("/home/placement/Downloads/rainfall in india 1901-2015.csv")

In [2]: data.describe()

Out[2]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	
count	4116.000000	4112.000000	4113.000000	4110.000000	4112.000000	4113.000000	4111.000000	4109.000000	4112.000000	4110.000000	4109.0
mean	1958.218659	18.957320	21.805325	27.359197	43.127432	85.745417	230.234444	347.214334	290.263497	197.361922	95.!
std	33.140898	33.585371	35.909488	46.959424	67.831168	123.234904	234.710758	269.539667	188.770477	135.408345	99.!
min	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.400000	0.000000	0.000000	0.100000	0.0
25%	1930.000000	0.600000	0.600000	1.000000	3.000000	8.600000	70.350000	175.600000	155.975000	100.525000	14.0
50%	1958.000000	6.000000	6.700000	7.800000	15.700000	36.600000	138.700000	284.800000	259.400000	173.900000	65.1
75%	1987.000000	22.200000	26.800000	31.300000	49.950000	97.200000	305.150000	418.400000	377.800000	265.800000	148.
max	2015.000000	583.700000	403.500000	605.600000	595.100000	1168.600000	1609.900000	2362.800000	1664.600000	1222.000000	948.:

## In [3]: data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4116 entries, 0 to 4115
Data columns (total 19 columns):
     Column
                  Non-Null Count Dtype
     _ _ _ _ _
     SUBDIVISION
                  4116 non-null
                                   obiect
                  4116 non-null
                                   int64
 1
     YEAR
 2
     JAN
                  4112 non-null
                                   float64
 3
     FEB
                  4113 non-null
                                   float64
 4
     MAR
                  4110 non-null
                                   float64
 5
     APR
                  4112 non-null
                                   float64
 6
     MAY
                  4113 non-null
                                   float64
 7
     JUN
                  4111 non-null
                                   float64
 8
     JUL
                  4109 non-null
                                   float64
 9
     AUG
                  4112 non-null
                                   float64
 10
     SEP
                  4110 non-null
                                   float64
     0CT
                  4109 non-null
                                   float64
 11
 12
     NOV
                  4105 non-null
                                   float64
 13
     DEC
                  4106 non-null
                                   float64
    ANNUAL
                  4090 non-null
                                   float64
 14
 15
    Jan-Feb
                  4110 non-null
                                   float64
    Mar-May
                  4107 non-null
                                   float64
 16
 17
    Jun-Sep
                  4106 non-null
                                   float64
 18 Oct-Dec
                  4103 non-null
                                   float64
dtypes: float64(17), int64(1), object(1)
memory usage: 611.1+ KB
```

data.groupby(['SUBDIVISION']).count() In [4]: Out[4]: Jan-Mar-Jun-Oct-JUN JUL AUG SEP OCT NOV DEC ANNUAL YEAR JAN FEB MAR APR MAY Feb Mav Sep Dec **SUBDIVISION ANDAMAN & NICOBAR ISLANDS** ARUNACHAL PRADESH **ASSAM & MEGHALAYA BIHAR CHHATTISGARH COASTAL ANDHRA PRADESH COASTAL KARNATAKA EAST MADHYA PRADESH** EAST RAJASTHAN **EAST UTTAR PRADESH GANGETIC WEST BENGAL GUJARAT REGION** HARYANA DELHI & **CHANDIGARH HIMACHAL PRADESH JAMMU & KASHMIR JHARKHAND KERALA KONKAN & GOA LAKSHADWEEP** MADHYA MAHARASHTRA MATATHWADA NAGA MANI MIZO TRIPURA 

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	ост	NOV	DEC	ANNUAL	Jan- Feb	Mar- May	Jun- Sep	Oct- Dec
SUBDIVISION																		
NORTH INTERIOR KARNATAKA	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
ORISSA	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
PUNJAB	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
RAYALSEEMA	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
SAURASHTRA & KUTCH	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
SOUTH INTERIOR KARNATAKA	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
SUB HIMALAYAN WEST BENGAL & SIKKIM	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
TAMIL NADU	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
TELANGANA	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
UTTARAKHAND	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
VIDARBHA	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
WEST MADHYA PRADESH	115	115	114	115	115	115	115	115	115	115	115	115	115	114	114	115	115	115
WEST RAJASTHAN	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
WEST UTTAR PRADESH	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115

```
In [6]: data.isna().sum()
Out[6]: SUBDIVISION
                         0
        YEAR
                         0
        JAN
                          4
        FEB
                          3
        MAR
                          6
        APR
                          4
        MAY
        JUN
        JUL
        AUG
        SEP
                          6
        0CT
                         7
        NOV
                        11
        DEC
                        10
        ANNUAL
                        26
        Jan-Feb
                         6
        Mar-May
                         9
        Jun-Sep
                        10
        Oct-Dec
                        13
        dtype: int64
In [7]: data1=data.loc[(data.YEAR<=2010)]</pre>
```

In [8]: data1

Out[8]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	ост	NOV	DEC	ANNUAL	Jan- Feb	Mar- May	Jun- Sep	
0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	388.5	558.2	33.6	3373.2	136.3	560.3	1696.3	g
1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2	197.2	359.0	160.5	3520.7	159.8	458.3	2185.9	7
2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0	181.2	284.4	225.0	2957.4	156.7	236.1	1874.0	6
3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4	222.2	308.7	40.1	3079.6	24.1	506.9	1977.6	5
4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0	260.7	25.4	344.7	2566.7	1.3	309.7	1624.9	6
															•••				
4106	LAKSHADWEEP	2006	20.1	0.0	33.0	0.3	327.9	286.9	172.3	150.7	318.5	119.1	158.9	10.9	1598.6	20.1	361.2	928.4	2
4107	LAKSHADWEEP	2007	2.5	4.2	0.2	22.2	166.2	573.4	427.4	294.7	457.5	256.1	47.6	109.6	2361.6	6.7	188.6	1753.0	4
4108	LAKSHADWEEP	2008	5.5	19.8	120.7	15.8	180.4	254.6	363.9	206.6	108.9	252.9	67.6	130.1	1726.8	25.3	316.9	934.0	4
4109	LAKSHADWEEP	2009	4.7	1.5	0.1	18.1	162.1	401.2	266.4	185.0	145.1	87.4	166.2	132.3	1570.1	6.2	180.3	997.7	3
4110	LAKSHADWEEP	2010	18.8	0.0	1.2	35.6	79.0	318.9	336.7	335.1	161.5	155.4	201.5	81.5	1725.2	18.8	115.8	1152.2	4

3936 rows × 19 columns

In [9]: data2=data.drop(['ANNUAL','Jan-Feb','Mar-May','Jun-Sep','Oct-Dec'],axis=1)

In [10]: data2

## Out[10]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	ОСТ	NOV	DEC
0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	388.5	558.2	33.6
1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2	197.2	359.0	160.5
2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0	181.2	284.4	225.0
3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4	222.2	308.7	40.1
4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0	260.7	25.4	344.7
4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2	117.4	184.3	14.9
4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8	145.9	12.4	8.8
4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.0	72.8	78.1	26.7
4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.2	169.2	59.0	62.3
4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	296.6	257.5	146.4	160.4	165.4	231.0	159.0

4116 rows × 14 columns

```
In [11]: data1['SUBDIVISION'].unique()
Out[11]: array(['ANDAMAN & NICOBAR ISLANDS', 'ARUNACHAL PRADESH',
```

```
'ASSAM & MEGHALAYA', 'NAGA MANI MIZO TRIPURA',
'SUB HIMALAYAN WEST BENGAL & SIKKIM', 'GANGETIC WEST BENGAL',
'ORISSA', 'JHARKHAND', 'BIHAR', 'EAST UTTAR PRADESH',
'WEST UTTAR PRADESH', 'UTTARAKHAND', 'HARYANA DELHI & CHANDIGARH',
'PUNJAB', 'HIMACHAL PRADESH', 'JAMMU & KASHMIR', 'WEST RAJASTHAN',
'EAST RAJASTHAN', 'WEST MADHYA PRADESH', 'EAST MADHYA PRADESH',
'GUJARAT REGION', 'SAURASHTRA & KUTCH', 'KONKAN & GOA',
'MADHYA MAHARASHTRA', 'MATATHWADA', 'VIDARBHA', 'CHHATTISGARH',
```

- 'COASTAL ANDHRA PRADESH', 'TELANGANA', 'RAYALSEEMA', 'TAMIL NADU',
- 'COASTAL KARNATAKA', 'NORTH INTERIOR KARNATAKA',
- 'SOUTH INTERIOR KARNATAKA', 'KERALA', 'LAKSHADWEEP'], dtype=object)

```
In [16]: | data2=data1.loc[(data1.SUBDIVISION== 'KERALA')]
In [17]: data2.isna().sum()
Out[17]: SUBDIVISION
                         0
         YEAR
                         0
         JAN
                         0
         FEB
                         0
         MAR
         APR
         MAY
         JUN
         JUL
         AUG
         SEP
         0CT
         NOV
         DEC
         ANNUAL
         Jan-Feb
         Mar-May
         Jun-Sep
         Oct-Dec
                         0
         dtype: int64
In [24]: data2['ANNUAL RAIN']=data2.apply(lambda row:row.JAN+row.FEB,axis=1)
In [20]: import warnings
         warnings.filterwarnings('ignore')
In [21]: data2['ANNUAL RAIN']=data2.apply(lambda row:row.JAN+row.FEB,axis=1)
```

In [22]: data2

Out[22]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	ОСТ	NOV	DEC	ANNUAL	Jan- Feb	Mar- May	Jun- Sep	Oc Di
3887	KERALA	1901	28.7	44.7	51.6	160.0	174.7	824.6	743.0	357.5	197.7	266.9	350.8	48.4	3248.6	73.4	386.2	2122.8	666
3888	KERALA	1902	6.7	2.6	57.3	83.9	134.5	390.9	1205.0	315.8	491.6	358.4	158.3	121.5	3326.6	9.3	275.7	2403.4	638
3889	KERALA	1903	3.2	18.6	3.1	83.6	249.7	558.6	1022.5	420.2	341.8	354.1	157.0	59.0	3271.2	21.7	336.3	2343.0	570
3890	KERALA	1904	23.7	3.0	32.2	71.5	235.7	1098.2	725.5	351.8	222.7	328.1	33.9	3.3	3129.7	26.7	339.4	2398.2	365
3891	KERALA	1905	1.2	22.3	9.4	105.9	263.3	850.2	520.5	293.6	217.2	383.5	74.4	0.2	2741.6	23.4	378.5	1881.5	458
3992	KERALA	2006	8.1	0.5	90.7	65.3	521.2	482.4	804.0	432.6	474.8	376.4	162.8	1.8	3420.6	8.6	677.2	2193.8	541
3993	KERALA	2007	0.5	5.6	7.3	138.5	192.7	705.9	966.3	489.6	526.7	357.2	87.4	11.9	3489.6	6.1	338.4	2688.5	456
3994	KERALA	2008	8.0	30.3	217.2	108.4	81.2	469.9	505.1	349.0	347.0	343.4	55.4	17.0	2524.5	31.1	406.7	1670.9	415
3995	KERALA	2009	3.3	1.5	62.6	69.0	191.6	438.2	924.9	269.3	326.5	205.2	274.4	44.2	2810.6	4.8	323.1	1958.9	523
3996	KERALA	2010	18.6	1.0	31.4	138.9	190.6	667.5	629.0	356.0	275.6	441.4	335.1	46.8	3131.8	19.6	360.9	1928.0	823

110 rows × 20 columns

In [25]: data2['ANNUAL RAIN']=data2.apply(lambda row:row.JAN+row.FEB+row.MAR+row.APR+row.MAY+row.JUN+row.JUL+row.AUG+

In [26]: data2

Out[26]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	ОСТ	NOV	DEC	ANNUAL	Jan- Feb	Mar- May	Jun- Sep	Oc Di
3887	KERALA	1901	28.7	44.7	51.6	160.0	174.7	824.6	743.0	357.5	197.7	266.9	350.8	48.4	3248.6	73.4	386.2	2122.8	666
3888	KERALA	1902	6.7	2.6	57.3	83.9	134.5	390.9	1205.0	315.8	491.6	358.4	158.3	121.5	3326.6	9.3	275.7	2403.4	638
3889	KERALA	1903	3.2	18.6	3.1	83.6	249.7	558.6	1022.5	420.2	341.8	354.1	157.0	59.0	3271.2	21.7	336.3	2343.0	570
3890	KERALA	1904	23.7	3.0	32.2	71.5	235.7	1098.2	725.5	351.8	222.7	328.1	33.9	3.3	3129.7	26.7	339.4	2398.2	365
3891	KERALA	1905	1.2	22.3	9.4	105.9	263.3	850.2	520.5	293.6	217.2	383.5	74.4	0.2	2741.6	23.4	378.5	1881.5	458
3992	KERALA	2006	8.1	0.5	90.7	65.3	521.2	482.4	804.0	432.6	474.8	376.4	162.8	1.8	3420.6	8.6	677.2	2193.8	541
3993	KERALA	2007	0.5	5.6	7.3	138.5	192.7	705.9	966.3	489.6	526.7	357.2	87.4	11.9	3489.6	6.1	338.4	2688.5	456
3994	KERALA	2008	8.0	30.3	217.2	108.4	81.2	469.9	505.1	349.0	347.0	343.4	55.4	17.0	2524.5	31.1	406.7	1670.9	415
3995	KERALA	2009	3.3	1.5	62.6	69.0	191.6	438.2	924.9	269.3	326.5	205.2	274.4	44.2	2810.6	4.8	323.1	1958.9	523
3996	KERALA	2010	18.6	1.0	31.4	138.9	190.6	667.5	629.0	356.0	275.6	441.4	335.1	46.8	3131.8	19.6	360.9	1928.0	823

110 rows × 20 columns

localhost:8888/notebooks/rainfall in india.ipynb

In [27]: cor=data.corr()
cor

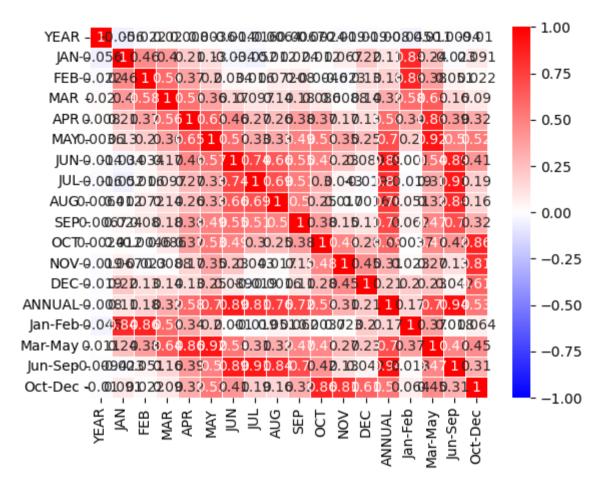
Out[27]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	ОСТ	NOV	DEC
YEAR	1.000000	-0.056235	-0.022144	0.020338	0.008007	0.003594	-0.013594	-0.016240	0.006442	-0.006670	0.002406	-0.018776	-0.019139
JAN	-0.056235	1.000000	0.456183	0.398502	0.209302	0.129622	-0.033725	-0.051642	0.011952	0.024289	0.012374	0.067281	0.219701
FEB	-0.022144	0.456183	1.000000	0.579819	0.367114	0.203062	0.033703	0.016235	0.072159	0.080148	-0.004581	-0.023413	0.132570
MAR	0.020338	0.398502	0.579819	1.000000	0.556856	0.362815	0.165857	0.097334	0.135071	0.178904	0.086187	0.008814	0.136328
APR	0.008007	0.209302	0.367114	0.556856	1.000000	0.650595	0.457091	0.268097	0.256168	0.382525	0.368886	0.165642	0.132892
MAY	0.003594	0.129622	0.203062	0.362815	0.650595	1.000000	0.567618	0.332283	0.329499	0.492378	0.529342	0.351931	0.250112
JUN	-0.013594	-0.033725	0.033703	0.165857	0.457091	0.567618	1.000000	0.741285	0.655142	0.551890	0.490393	0.229718	0.088782
JUL	-0.016240	-0.051642	0.016235	0.097334	0.268097	0.332283	0.741285	1.000000	0.686662	0.513067	0.299221	0.042671	-0.019427
AUG	0.006442	0.011952	0.072159	0.135071	0.256168	0.329499	0.655142	0.686662	1.000000	0.497037	0.250600	0.017488	0.001648
SEP	-0.006670	0.024289	0.080148	0.178904	0.382525	0.492378	0.551890	0.513067	0.497037	1.000000	0.384138	0.153465	0.109457
ОСТ	0.002406	0.012374	-0.004581	0.086187	0.368886	0.529342	0.490393	0.299221	0.250600	0.384138	1.000000	0.477503	0.281172
NOV	-0.018776	0.067281	-0.023413	0.008814	0.165642	0.351931	0.229718	0.042671	0.017488	0.153465	0.477503	1.000000	0.451407
DEC	-0.019139	0.219701	0.132570	0.136328	0.132892	0.250112	0.088782	-0.019427	0.001648	0.109457	0.281172	0.451407	1.000000
ANNUAL	-0.008044	0.105696	0.181563	0.322199	0.577573	0.698013	0.891303	0.812279	0.759304	0.715135	0.587065	0.308768	0.207176
Jan-Feb	-0.044653	0.842390	0.863815	0.576366	0.340841	0.196168	0.001016	-0.019157	0.050918	0.062131	0.003743	0.022885	0.204848
Mar-May	0.010637	0.242256	0.382620	0.642294	0.864172	0.915019	0.538562	0.313726	0.318347	0.470032	0.468048	0.272268	0.228473
Jun-Sep	-0.009418	-0.022748	0.051066	0.162055	0.394859	0.496164	0.893968	0.907723	0.840352	0.701980	0.416350	0.126338	0.042440
Oct-Dec	-0.010155	0.090932	0.021878	0.090108	0.321407	0.523684	0.409050	0.190400	0.156293	0.319832	0.862761	0.808798	0.606658

localhost:8888/notebooks/rainfall in india.ipynb

```
In [30]: import seaborn as sns
         sns.heatmap(cor,vmax=1,vmin=-1,annot=True,linewidth=.5,cmap='bwr')
```

Out[30]: <Axes: >



```
In [32]: data2=data2.drop(['SUBDIVISION','JAN','FEB','MAR','APR','MAY','JUN','JUL','AUG','SEP','OCT','NOV','DEC','ANN
```

In [33]: data2

Out[33]:

	YEAR	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec	ANNUAL RAIN
3887	1901	73.4	386.2	2122.8	666.1	3248.6
3888	1902	9.3	275.7	2403.4	638.2	3326.5
3889	1903	21.7	336.3	2343.0	570.1	3271.4
3890	1904	26.7	339.4	2398.2	365.3	3129.6
3891	1905	23.4	378.5	1881.5	458.1	2741.7
3992	2006	8.6	677.2	2193.8	541.0	3420.6
3993	2007	6.1	338.4	2688.5	456.5	3489.6
3994	2008	31.1	406.7	1670.9	415.7	2524.7
3995	2009	4.8	323.1	1958.9	523.8	2810.7
3996	2010	19.6	360.9	1928.0	823.3	3131.9

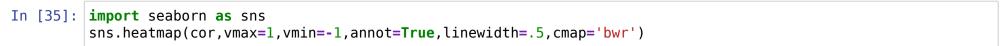
110 rows × 6 columns

In [34]: cor=data2.corr()

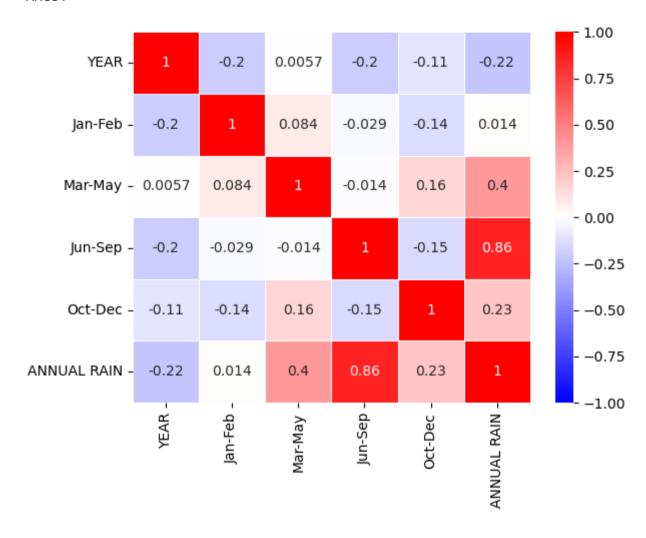
cor

Out[34]:

	YEAR	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec	ANNUAL RAIN
YEAR	1.000000	-0.195356	0.005666	-0.200507	-0.106562	-0.224014
Jan-Feb	-0.195356	1.000000	0.083762	-0.028765	-0.137987	0.014396
Mar-May	0.005666	0.083762	1.000000	-0.014441	0.157126	0.402308
Jun-Sep	-0.200507	-0.028765	-0.014441	1.000000	-0.145365	0.859485
Oct-Dec	-0.106562	-0.137987	0.157126	-0.145365	1.000000	0.226826
ANNUAL RAIN	-0.224014	0.014396	0.402308	0.859485	0.226826	1.000000



Out[35]: <Axes: >



In [ ]: