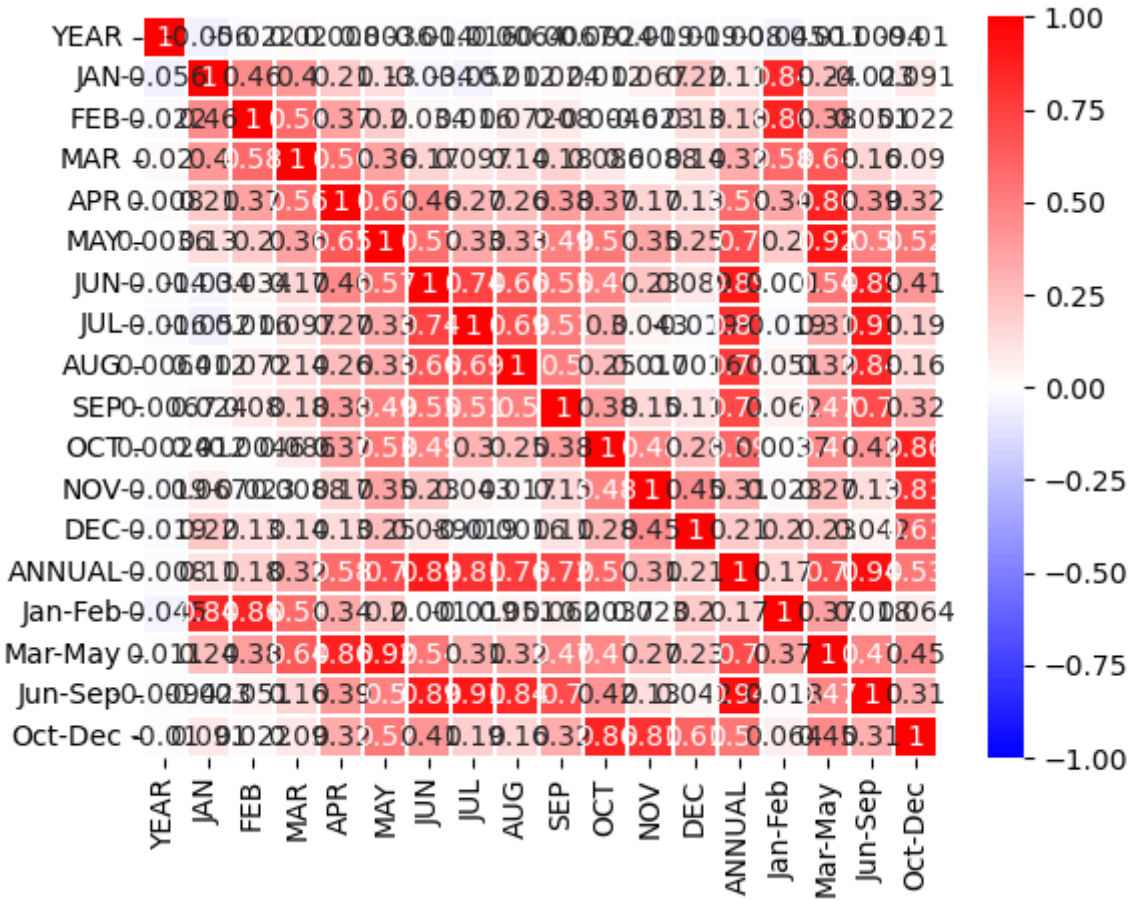
In [1]:
```
import pandas as r
d=r.read_csv("/home/placement/Downloads/rainfall in india 1901-2015.csv")
import warnings
warnings.filterwarnings("ignore")
```

In [2]:
```
#This function is to find correlation
cor=d.corr()
```

Out[2]: <Axes: >



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

In [3]:
```
#loc[] is used to retrieve the group of rows and columns by labels
d3=d.loc[(d.SUBDIVISION=='LAKSHADWEEP')]
```

In [4]:
```
'''We can count the NaN values in Pandas DataFrame using the isna() function
and with the sum() function '''
d3.isna().sum()
```

Out[4]:
```
SUBDIVISION     0
YEAR            0
JAN             2
FEB             1
MAR             2
APR             2
MAY             2
JUN             2
JUL             3
AUG             2
SEP             3
OCT             3
NOV             6
DEC             4
ANNUAL         11
Jan-Feb         3
Mar-May         4
Jun-Sep         4
Oct-Dec         6
dtype: int64
```

In [5]:
```
#filling the NaN
d3=d3.fillna(d3.mean())
```

In [6]:
```
#The data after filling the NaN values
d3.head(200)
```

Out[6]:

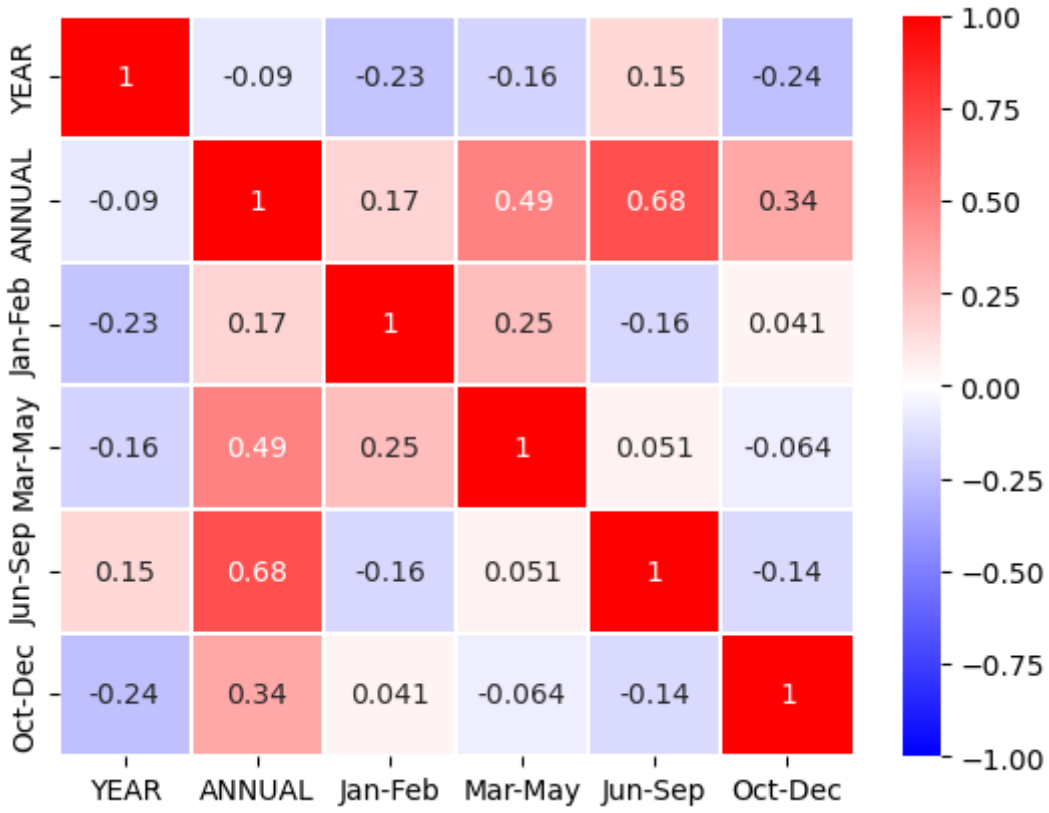
	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Ma Ma
4002	LAKSHADWEEP	1901	22.6	86.4	114.8	263.8	37.3	459.0	0.0	0.0	46.7	183.7	229.900000	15.000000	1459.200000	109.0	415
4003	LAKSHADWEEP	1902	99.3	9.6	32.6	40.4	179.1	374.2	413.3	170.0	214.3	384.2	192.800000	49.000000	2158.800000	108.9	252
4004	LAKSHADWEEP	1903	63.5	95.0	0.0	29.5	144.1	212.4	261.8	202.0	292.1	79.1	124.840741	60.810909	1590.886408	158.5	173
4005	LAKSHADWEEP	1904	0.0	0.0	13.5	13.2	143.3	261.3	256.0	38.9	219.9	153.6	8.300000	68.900000	1176.900000	0.0	170
4006	LAKSHADWEEP	1905	62.4	0.0	0.0	0.0	166.7	400.7	68.7	377.5	107.5	232.1	159.300000	0.000000	1574.900000	62.4	166
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2	117.4	184.300000	14.900000	1533.700000	7.9	196
4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8	145.9	12.400000	8.800000	1405.500000	19.3	99
4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.0	72.8	78.100000	26.700000	1426.300000	60.6	131
4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.2	169.2	59.000000	62.300000	1395.000000	69.3	76
4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	296.6	257.5	146.4	160.4	165.4	231.000000	159.000000	1642.900000	2.7	223

114 rows × 19 columns

In [7]:
```
#Removing the columns in the dataframe
d3=d3.drop(['JAN','FEB','MAR','APR','MAY','JUN','JUL','AUG','SEP','OCT','NOV','DEC'],axis=1)
```

In [8]:
```
#This function is to find correlation
cor=d3.corr()
#This create a heatmap
import seaborn as sns
sns.heatmap(cor,vmax=1,vmin=-1,annot=True,linewidth=.1,cmap='bwr')
```

Out[8]: <Axes: >



In [9]:
```
#loc[] is used to retrieve the group of rows and columns by labels
d1=d.loc[(d.YEAR<=2010)]
d1
```

Out[9]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	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	9
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 [10]:
```
d1=d1.drop(['ANNUAL','Jan-Feb','Mar-May','Jun-Sep','Oct-Dec'],axis=1)
d1
```

Out[10]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
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
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
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
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
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

3936 rows × 14 columns

In [11]:
```
#To get unique values in a column
d1['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 [12]:
```
#Filling the NaN values
d=d.fillna(d.mean())
```

In [13]:
```
d.tail(100)
```

Out[13]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
4016	LAKSHADWEEP	1915	0.00000	0.000000	22.900000	50.500000	143.300000	183.400000	461.600000	76.000000	222.100000	151.600000
4017	LAKSHADWEEP	1916	0.30000	13.200000	32.000000	28.700000	130.500000	327.600000	445.900000	31.800000	275.400000	85.200000
4018	LAKSHADWEEP	1917	18.95732	68.600000	27.359197	43.127432	85.745417	230.234444	347.214334	290.263497	197.361922	95.507009
4019	LAKSHADWEEP	1919	74.40000	1.300000	39.400000	25.600000	249.800000	352.100000	367.400000	175.600000	196.900000	181.000000
4020	LAKSHADWEEP	1920	82.60000	21.805325	35.600000	49.400000	203.100000	530.000000	134.200000	116.900000	247.600000	258.800000
...	...	...	...	...	...	...	...	...	...	...	...	...	...
4111	LAKSHADWEEP	2011	5.10000	2.800000	3.100000	85.900000	107.200000	153.600000	350.200000	254.000000	255.200000	117.400000
4112	LAKSHADWEEP	2012	19.20000	0.100000	1.600000	76.800000	21.200000	327.000000	231.500000	381.200000	179.800000	145.900000
4113	LAKSHADWEEP	2013	26.20000	34.400000	37.500000	5.300000	88.300000	426.200000	296.400000	154.400000	180.000000	72.800000
4114	LAKSHADWEEP	2014	53.20000	16.100000	4.400000	14.900000	57.400000	244.100000	116.100000	466.100000	132.200000	169.200000
4115	LAKSHADWEEP	2015	2.20000	0.500000	3.700000	87.100000	133.100000	296.600000	257.500000	146.400000	160.400000	165.400000

100 rows × 19 columns

In [14]:
```
d1.groupby('SUBDIVISION').count()
```

Out[14]:

SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
ANDAMAN & NICOBAR ISLANDS	105	105	105	103	103	104	103	103	103	102	103	103	102
ARUNACHAL PRADESH	92	91	91	90	92	92	91	91	92	92	90	90	90
ASSAM & MEGHALAYA	110	110	110	110	110	110	110	110	110	110	110	110	110
BIHAR	110	110	110	110	110	110	110	110	110	110	110	110	110
CHHATTISGARH	110	110	110	110	110	110	110	110	110	110	110	110	110
COASTAL ANDHRA PRADESH	110	110	110	110	110	110	110	110	110	110	110	110	110
COASTAL KARNATAKA	110	110	110	110	110	110	110	110	110	110	110	110	110
EAST MADHYA PRADESH	110	110	110	110	110	110	110	110	110	110	110	110	110
EAST RAJASTHAN	110	110	110	110	110	110	110	110	110	110	110	110	110
EAST UTTAR PRADESH	110	110	110	110	110	110	110	110	110	110	110	110	110
GANGETIC WEST BENGAL	110	110	110	110	110	110	110	110	110	110	110	110	110
GUJARAT REGION	110	110	110	110	110	110	110	110	110	110	110	110	110
HARYANA DELHI & CHANDIGARH	110	110	110	110	110	110	110	110	110	110	110	110	110
HIMACHAL PRADESH	110	110	110	110	110	110	110	110	110	110	110	110	110
JAMMU & KASHMIR	110	110	110	110	110	110	110	109	110	110	110	109	109
JHARKHAND	110	110	110	110	110	110	110	110	110	110	110	110	110
KERALA	110	110	110	110	110	110	110	110	110	110	110	110	110
KONKAN & GOA	110	110	110	110	110	110	110	110	110	110	110	110	110
LAKSHADWEEP	109	107	108	107	107	107	107	106	107	106	106	103	105
MADHYA MAHARASHTRA	110	110	110	110	110	110	110	110	110	110	110	110	110
MATATHWADA	110	110	110	110	110	110	110	110	110	110	110	110	110
NAGA MANI MIZO TRIPURA	110	110	110	110	110	110	110	110	110	110	110	110	110
NORTH INTERIOR KARNATAKA	110	110	110	110	110	110	110	110	110	110	110	110	110

SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
ORISSA	110	110	110	110	110	110	110	110	110	110	110	110	110
PUNJAB	110	110	110	110	110	110	110	110	110	110	110	110	110
RAYALSEEMA	110	110	110	110	110	110	110	110	110	110	110	110	110
SAURASHTRA & KUTCH	110	110	110	110	110	110	110	110	110	110	110	110	110
SOUTH INTERIOR KARNATAKA	110	110	110	110	110	110	110	110	110	110	110	110	110
SUB HIMALAYAN WEST BENGAL & SIKKIM	110	110	110	110	110	110	110	110	110	110	110	110	110
TAMIL NADU	110	110	110	110	110	110	110	110	110	110	110	110	110
TELANGANA	110	110	110	110	110	110	110	110	110	110	110	110	110
UTTARAKHAND	110	110	110	110	110	110	110	110	110	110	110	110	110
VIDARBHA	110	110	110	110	110	110	110	110	110	110	110	110	110
WEST MADHYA PRADESH	110	110	109	110	110	110	110	110	110	110	110	110	110
WEST RAJASTHAN	110	110	110	110	110	110	110	110	110	110	110	110	110
WEST UTTAR PRADESH	110	110	110	110	110	110	110	110	110	110	110	110	110

In [15]:
```
#TO ADD COLUMNS and create new column
d1['NEM']=d1.apply(lambda row:row.OCT +row.NOV +row.DEC,axis=1) #TO ADD NEW COLUMNS
d1
```

Out[15]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	NEM
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	980.3
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	716.7
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	690.6
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	571.0
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	630.8
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
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	288.9
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	413.3
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	450.6
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	385.9
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	438.4

3936 rows × 15 columns

In [ ]: