In [1]:
```
import pandas as pd
```

In [2]:
```
data=pd.read_csv("/home/placement/Desktop/venkatesh/rainfall in india 1901-2015.csv")
```

In [3]:
```
data.groupby(['SUBDIVISION']).count()
```

Out[3]:

SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec
ANDAMAN & NICOBAR ISLANDS	110	110	110	108	108	109	108	108	108	107	108	108	107	104	110	107	107	107
ARUNACHAL PRADESH	97	96	96	95	97	97	96	96	97	97	95	95	95	91	96	95	95	94
ASSAM & MEGHALAYA	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
BIHAR	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
CHHATTISGARH	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
COASTAL ANDHRA PRADESH	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
COASTAL KARNATAKA	115	114	115	115	115	115	115	115	115	115	115	115	115	114	114	115	115	115
EAST MADHYA PRADESH	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
EAST RAJASTHAN	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
EAST UTTAR PRADESH	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
GANGETIC WEST BENGAL	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
GUJARAT REGION	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
HARYANA DELHI & CHANDIGARH	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
HIMACHAL PRADESH	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
JAMMU & KASHMIR	115	115	115	115	115	115	115	114	115	115	115	114	114	114	115	115	114	114
JHARKHAND	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
KERALA	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
KONKAN & GOA	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
LAKSHADWEEP	114	112	113	112	112	112	112	111	112	111	111	108	110	103	111	110	110	108
MADHYA MAHARASHTRA	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
MATATHWADA	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115

SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec
NAGA MANI MIZO TRIPURA	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
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 [4]:
```
data['SUBDIVISION'].unique()
```

Out[4]:
```
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 [5]:
```
data.isna().sum()
```

Out[5]:
```
SUBDIVISION     0
YEAR            0
JAN             4
FEB             3
MAR             6
APR             4
MAY             3
JUN             5
JUL             7
AUG             4
SEP             6
OCT             7
NOV            11
DEC            10
ANNUAL         26
Jan-Feb         6
Mar-May         9
Jun-Sep        10
Oct-Dec        13
dtype: int64
```

In [6]:
```
data1=data.drop(['ANNUAL','Jan-Feb','Mar-May','Jun-Sep','Oct-Dec'],axis=1)
```

In [7]:
```
data1
```

Out[7]:

	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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
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 [8]:
```
data2=data1.loc[(data.SUBDIVISION=='TELANGANA')]
```

In [9]:
```
data2
```

Out[9]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
3197	TELANGANA	1901	6.9	41.8	7.8	45.2	22.0	123.6	237.8	177.2	77.7	75.5	12.2	0.0
3198	TELANGANA	1902	0.0	0.0	0.2	10.7	7.3	52.4	146.3	142.8	190.5	41.7	31.2	7.3
3199	TELANGANA	1903	12.9	4.6	0.0	9.9	40.7	99.2	505.2	246.7	191.9	155.8	15.5	1.1
3200	TELANGANA	1904	0.0	0.0	10.8	0.8	14.7	104.2	139.5	50.0	162.3	44.4	0.0	0.0
3201	TELANGANA	1905	0.0	4.3	12.8	27.6	32.2	129.5	82.4	237.3	179.1	19.6	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3307	TELANGANA	2011	0.0	11.9	2.6	25.6	9.3	83.9	268.2	225.9	107.6	13.9	4.2	0.0
3308	TELANGANA	2012	6.7	0.0	0.2	14.0	8.4	124.4	300.3	229.9	202.4	83.6	38.7	0.0
3309	TELANGANA	2013	2.4	29.0	0.2	24.4	8.5	213.4	453.8	230.6	161.4	205.9	16.4	2.7
3310	TELANGANA	2014	0.2	2.9	58.3	10.3	73.3	62.3	146.0	205.2	146.8	29.6	10.8	0.7
3311	TELANGANA	2015	17.5	0.0	43.0	65.7	23.3	266.9	104.4	160.5	158.3	15.6	0.3	1.7

115 rows × 14 columns

In [10]:
```
data3=data1.loc[(data.SUBDIVISION=='TELANGANA')]
```

In [11]:
```
data3
```

Out[11]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
3197	TELANGANA	1901	6.9	41.8	7.8	45.2	22.0	123.6	237.8	177.2	77.7	75.5	12.2	0.0
3198	TELANGANA	1902	0.0	0.0	0.2	10.7	7.3	52.4	146.3	142.8	190.5	41.7	31.2	7.3
3199	TELANGANA	1903	12.9	4.6	0.0	9.9	40.7	99.2	505.2	246.7	191.9	155.8	15.5	1.1
3200	TELANGANA	1904	0.0	0.0	10.8	0.8	14.7	104.2	139.5	50.0	162.3	44.4	0.0	0.0
3201	TELANGANA	1905	0.0	4.3	12.8	27.6	32.2	129.5	82.4	237.3	179.1	19.6	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3307	TELANGANA	2011	0.0	11.9	2.6	25.6	9.3	83.9	268.2	225.9	107.6	13.9	4.2	0.0
3308	TELANGANA	2012	6.7	0.0	0.2	14.0	8.4	124.4	300.3	229.9	202.4	83.6	38.7	0.0
3309	TELANGANA	2013	2.4	29.0	0.2	24.4	8.5	213.4	453.8	230.6	161.4	205.9	16.4	2.7
3310	TELANGANA	2014	0.2	2.9	58.3	10.3	73.3	62.3	146.0	205.2	146.8	29.6	10.8	0.7
3311	TELANGANA	2015	17.5	0.0	43.0	65.7	23.3	266.9	104.4	160.5	158.3	15.6	0.3	1.7

115 rows × 14 columns

In [12]:
```
data3['ANNUAL RAIN']=data3.apply(lambda row:row.JAN+row.FEB+row.MAR+row.APR+row.MAY+row.JUN+row.JUL+row.AUG+
```

```
/tmp/ipykernel_12853/57770726.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html
#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#retu
rning-a-view-versus-a-copy)
  data3['ANNUAL RAIN']=data3.apply(lambda row:row.JAN+row.FEB+row.MAR+row.APR+row.MAY+row.JUN+row.JUL+row.A
UG+row.SEP+row.OCT+row.NOV+row.DEC,axis=1)
```

In [13]:
```
data3
```

Out[13]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL RAIN
3197	TELANGANA	1901	6.9	41.8	7.8	45.2	22.0	123.6	237.8	177.2	77.7	75.5	12.2	0.0	827.7
3198	TELANGANA	1902	0.0	0.0	0.2	10.7	7.3	52.4	146.3	142.8	190.5	41.7	31.2	7.3	630.4
3199	TELANGANA	1903	12.9	4.6	0.0	9.9	40.7	99.2	505.2	246.7	191.9	155.8	15.5	1.1	1283.5
3200	TELANGANA	1904	0.0	0.0	10.8	0.8	14.7	104.2	139.5	50.0	162.3	44.4	0.0	0.0	526.7
3201	TELANGANA	1905	0.0	4.3	12.8	27.6	32.2	129.5	82.4	237.3	179.1	19.6	0.0	0.0	724.8
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3307	TELANGANA	2011	0.0	11.9	2.6	25.6	9.3	83.9	268.2	225.9	107.6	13.9	4.2	0.0	753.1
3308	TELANGANA	2012	6.7	0.0	0.2	14.0	8.4	124.4	300.3	229.9	202.4	83.6	38.7	0.0	1008.6
3309	TELANGANA	2013	2.4	29.0	0.2	24.4	8.5	213.4	453.8	230.6	161.4	205.9	16.4	2.7	1348.7
3310	TELANGANA	2014	0.2	2.9	58.3	10.3	73.3	62.3	146.0	205.2	146.8	29.6	10.8	0.7	746.4
3311	TELANGANA	2015	17.5	0.0	43.0	65.7	23.3	266.9	104.4	160.5	158.3	15.6	0.3	1.7	857.2

115 rows × 15 columns

In [14]:
```
data4=data3.drop(['SUBDIVISION'],axis=1)
```

In [15]:
```
data4
```

Out[15]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL RAIN
3197	1901	6.9	41.8	7.8	45.2	22.0	123.6	237.8	177.2	77.7	75.5	12.2	0.0	827.7
3198	1902	0.0	0.0	0.2	10.7	7.3	52.4	146.3	142.8	190.5	41.7	31.2	7.3	630.4
3199	1903	12.9	4.6	0.0	9.9	40.7	99.2	505.2	246.7	191.9	155.8	15.5	1.1	1283.5
3200	1904	0.0	0.0	10.8	0.8	14.7	104.2	139.5	50.0	162.3	44.4	0.0	0.0	526.7
3201	1905	0.0	4.3	12.8	27.6	32.2	129.5	82.4	237.3	179.1	19.6	0.0	0.0	724.8
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3307	2011	0.0	11.9	2.6	25.6	9.3	83.9	268.2	225.9	107.6	13.9	4.2	0.0	753.1
3308	2012	6.7	0.0	0.2	14.0	8.4	124.4	300.3	229.9	202.4	83.6	38.7	0.0	1008.6
3309	2013	2.4	29.0	0.2	24.4	8.5	213.4	453.8	230.6	161.4	205.9	16.4	2.7	1348.7
3310	2014	0.2	2.9	58.3	10.3	73.3	62.3	146.0	205.2	146.8	29.6	10.8	0.7	746.4
3311	2015	17.5	0.0	43.0	65.7	23.3	266.9	104.4	160.5	158.3	15.6	0.3	1.7	857.2

115 rows × 14 columns

In [16]:
```
cor=data4.corr()
cor
```

Out[16]:

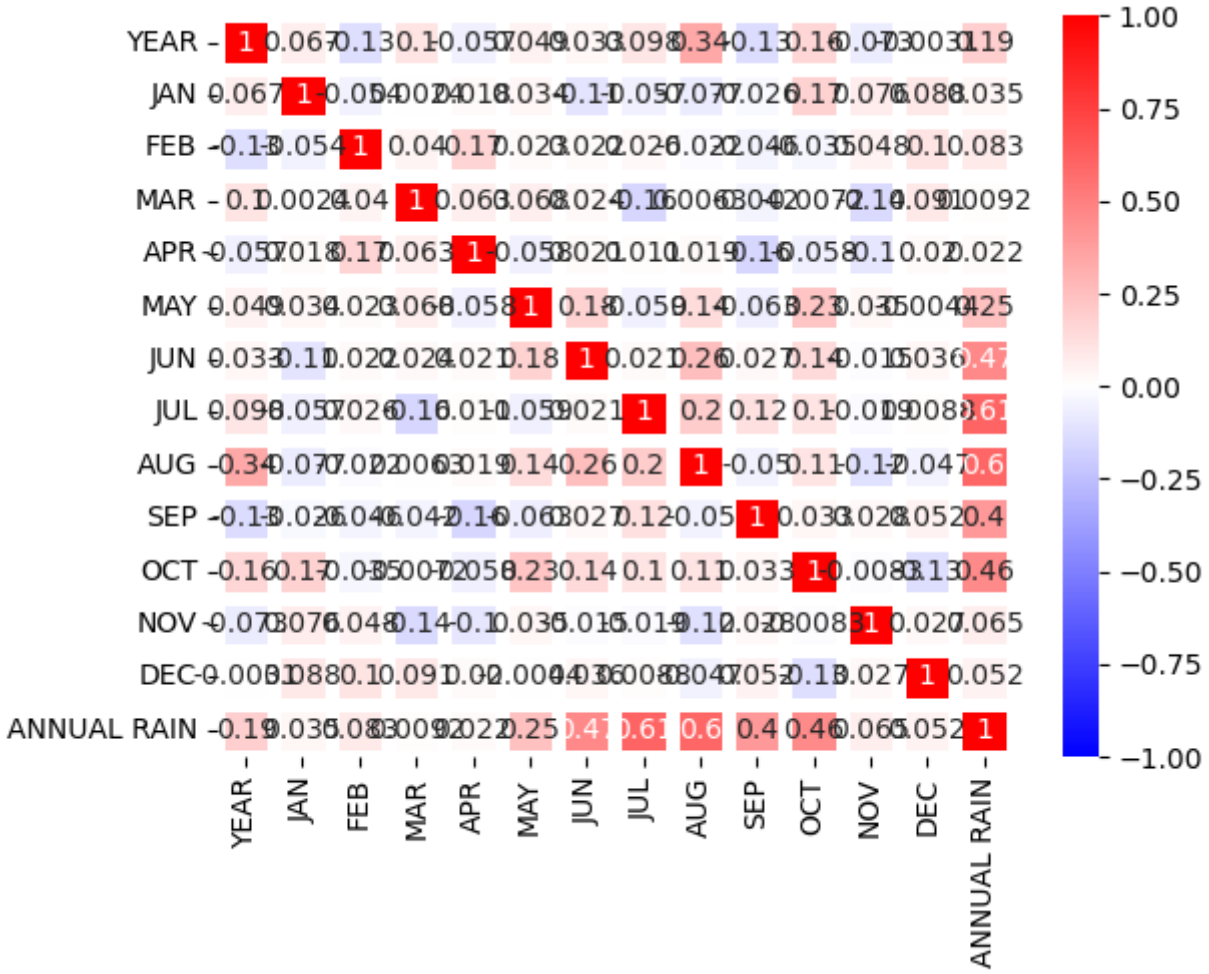
	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	D
YEAR	1.000000	0.066679	-0.127789	0.103372	-0.057459	0.049107	0.032522	0.097739	0.343648	-0.134013	0.156396	-0.072673	-0.003
JAN	0.066679	1.000000	-0.054219	0.002439	0.017525	0.033719	-0.107093	-0.056891	-0.076973	-0.026067	0.174812	0.076095	0.088
FEB	-0.127789	-0.054219	1.000000	0.039952	0.166871	0.022623	0.022467	0.026066	-0.021750	-0.045791	-0.035158	0.048416	0.101
MAR	0.103372	0.002439	0.039952	1.000000	0.062920	0.068031	0.024190	-0.162996	0.006257	-0.041883	-0.007192	-0.142749	0.091
APR	-0.057459	0.017525	0.166871	0.062920	1.000000	-0.057561	0.021496	0.011166	0.019299	-0.159302	-0.058436	-0.100563	0.020
MAY	0.049107	0.033719	0.022623	0.068031	-0.057561	1.000000	0.179190	-0.059126	0.137449	-0.063313	0.229135	0.034746	-0.004
JUN	0.032522	-0.107093	0.022467	0.024190	0.021496	0.179190	1.000000	0.021370	0.255568	0.026901	0.141178	-0.015491	0.035
JUL	0.097739	-0.056891	0.026066	-0.162996	0.011166	-0.059126	0.021370	1.000000	0.201899	0.123573	0.103609	-0.018848	0.008
AUG	0.343648	-0.076973	-0.021750	0.006257	0.019299	0.137449	0.255568	0.201899	1.000000	-0.049929	0.108571	-0.119876	-0.047
SEP	-0.134013	-0.026067	-0.045791	-0.041883	-0.159302	-0.063313	0.026901	0.123573	-0.049929	1.000000	0.032922	0.027962	0.051
OCT	0.156396	0.174812	-0.035158	-0.007192	-0.058436	0.229135	0.141178	0.103609	0.108571	0.032922	1.000000	-0.008315	-0.130
NOV	-0.072673	0.076095	0.048416	-0.142749	-0.100563	0.034746	-0.015491	-0.018848	-0.119876	0.027962	-0.008315	1.000000	0.026
DEC	-0.003100	0.088278	0.101628	0.091271	0.020066	-0.004376	0.035836	0.008781	-0.047230	0.051672	-0.130792	0.026896	1.000
ANNUAL RAIN	0.192611	0.035049	0.082843	0.009230	0.021805	0.248903	0.468318	0.609108	0.596173	0.400333	0.459548	0.064896	0.051

In [ ]:
```

```

In [17]:
```
import seaborn as sns
sns.heatmap(cor,vmax=1,vmin=-1,annot=True,linewidths=5,cmap='bwr')
```

Out[17]: <Axes: >



In [18]:
```
data4['SWM']=data4.apply(lambda row:row.JUN+row.JUL+row.AUG+row.SEP,axis=1)
```

In [19]:
```
data4['NEM']=data4.apply(lambda row:row.OCT+row.NOV+row.DEC,axis=1)#data2['NEM']=data2['OCT']+data2['NOV']+d
```

In [20]:
```
data4
```

Out[20]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL RAIN	SWM	NEM
3197	1901	6.9	41.8	7.8	45.2	22.0	123.6	237.8	177.2	77.7	75.5	12.2	0.0	827.7	616.3	87.7
3198	1902	0.0	0.0	0.2	10.7	7.3	52.4	146.3	142.8	190.5	41.7	31.2	7.3	630.4	532.0	80.2
3199	1903	12.9	4.6	0.0	9.9	40.7	99.2	505.2	246.7	191.9	155.8	15.5	1.1	1283.5	1043.0	172.4
3200	1904	0.0	0.0	10.8	0.8	14.7	104.2	139.5	50.0	162.3	44.4	0.0	0.0	526.7	456.0	44.4
3201	1905	0.0	4.3	12.8	27.6	32.2	129.5	82.4	237.3	179.1	19.6	0.0	0.0	724.8	628.3	19.6
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3307	2011	0.0	11.9	2.6	25.6	9.3	83.9	268.2	225.9	107.6	13.9	4.2	0.0	753.1	685.6	18.1
3308	2012	6.7	0.0	0.2	14.0	8.4	124.4	300.3	229.9	202.4	83.6	38.7	0.0	1008.6	857.0	122.3
3309	2013	2.4	29.0	0.2	24.4	8.5	213.4	453.8	230.6	161.4	205.9	16.4	2.7	1348.7	1059.2	225.0
3310	2014	0.2	2.9	58.3	10.3	73.3	62.3	146.0	205.2	146.8	29.6	10.8	0.7	746.4	560.3	41.1
3311	2015	17.5	0.0	43.0	65.7	23.3	266.9	104.4	160.5	158.3	15.6	0.3	1.7	857.2	690.1	17.6

115 rows × 16 columns

In [21]:
```
data5=data4.drop(['YEAR','JAN','FEB','MAR','APR','MAY','JUN','JUL','AUG','SEP','OCT','NOV','DEC'],axis=1)
```

In [22]:
```
data5
```

Out[22]:

	ANNUAL RAIN	SWM	NEM
3197	827.7	616.3	87.7
3198	630.4	532.0	80.2
3199	1283.5	1043.0	172.4
3200	526.7	456.0	44.4
3201	724.8	628.3	19.6
...	...	...	...
3307	753.1	685.6	18.1
3308	1008.6	857.0	122.3
3309	1348.7	1059.2	225.0
3310	746.4	560.3	41.1
3311	857.2	690.1	17.6

115 rows × 3 columns

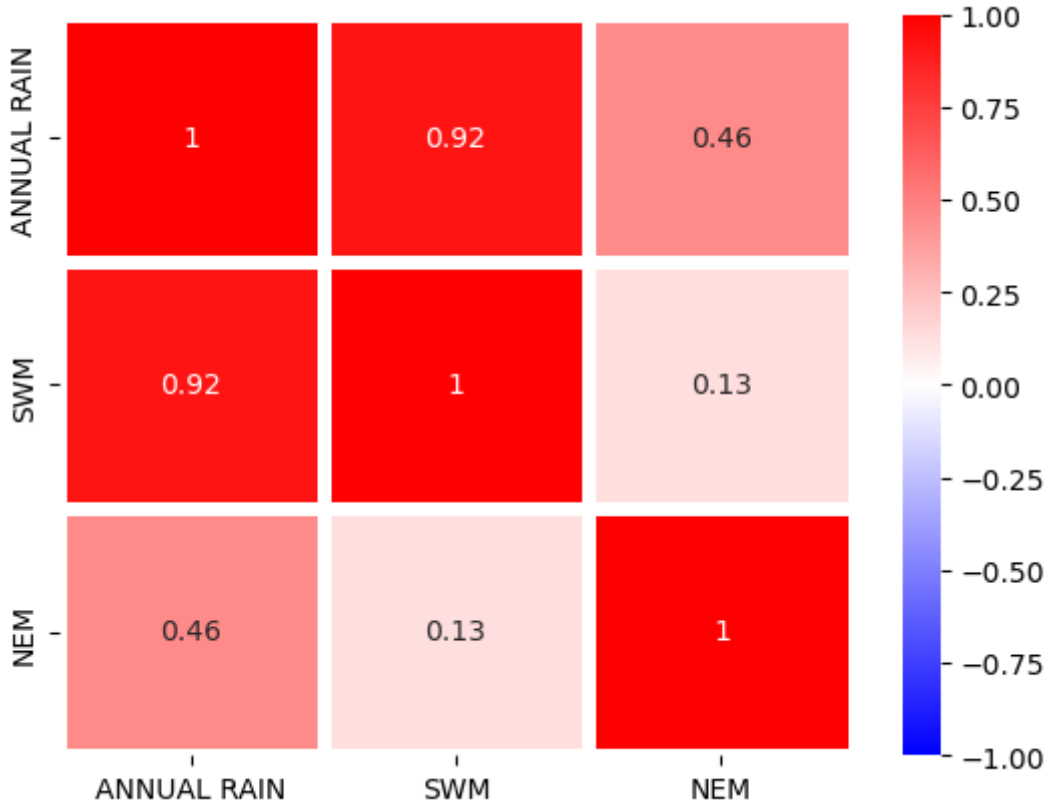
In [23]:
```
corel=data5.corr()
corel
```

Out[23]:

	ANNUAL RAIN	SWM	NEM
ANNUAL RAIN	1.000000	0.917503	0.458665
SWM	0.917503	1.000000	0.129536
NEM	0.458665	0.129536	1.000000

In [24]:
```
import seaborn as sns
sns.heatmap(corel,vmax=1,vmin=-1,annot=True,linewidths=5,cmap='bwr')
```

Out[24]: <Axes: >



In [ ]: