In [73]:
```
import pandas as pd
```

In [74]:
```
data=pd.read_csv("/home/placement/Desktop/prasanna/rainfall in india 1901-2015.crsv")#to read the file
```

In [75]:
```
data.describe()
```

Out[75]:

	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.(
mean	1958.218659	18.957320	21.805325	27.359197	43.127432	85.745417	230.234444	347.214334	290.263497	197.361922	95.5
std	33.140898	33.585371	35.909488	46.959424	67.831168	123.234904	234.710758	269.539667	188.770477	135.408345	99.5
min	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.400000	0.000000	0.000000	0.100000	0.(
25%	1930.000000	0.600000	0.600000	1.000000	3.000000	8.600000	70.350000	175.600000	155.975000	100.525000	14.(
50%	1958.000000	6.000000	6.700000	7.800000	15.700000	36.600000	138.700000	284.800000	259.400000	173.900000	65.2
75%	1987.000000	22.200000	26.800000	31.300000	49.950000	97.200000	305.150000	418.400000	377.800000	265.800000	148.4
max	2015.000000	583.700000	403.500000	605.600000	595.100000	1168.600000	1609.900000	2362.800000	1664.600000	1222.000000	948.3

In [76]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4116 entries, 0 to 4115
Data columns (total 19 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   SUBDIVISION  4116 non-null   object 
 1   YEAR         4116 non-null   int64  
 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
 11  OCT          4109 non-null   float64
 12  NOV          4105 non-null   float64
 13  DEC          4106 non-null   float64
 14  ANNUAL       4090 non-null   float64
 15  Jan-Feb      4110 non-null   float64
 16  Mar-May      4107 non-null   float64
 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
```

In [77]:
```
data.groupby(['SUBDIVISION']).count()
```

Out[77]:

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

In [78]:
```
data.isna().sum()
```

Out[78]:
```
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 [79]:
```
data1=data.loc[(data.YEAR<=2010)]
```

In [80]:
```
data1
```

Out[80]:

	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 [81]:
```
data2=data.drop(['ANNUAL','Jan-Feb','Mar-May','Jun-Sep','Oct-Dec'],axis=1)
```

In [82]:
```
data2
```

Out[82]:

	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 [83]:
```
data2['SUBDIVISION'].unique()
```

Out[83]:
```
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 [84]:
```
data3=data2.loc[(data2.SUBDIVISION=="COASTAL ANDHRA PRADESH")]
```

In [85]:
```
data3
```

Out[85]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
3082	COASTAL ANDHRA PRADESH	1901	18.8	80.9	7.2	28.7	68.7	77.7	113.0	133.7	125.3	173.4	164.8	1.5
3083	COASTAL ANDHRA PRADESH	1902	2.0	0.0	2.8	23.9	37.6	72.6	144.5	236.1	204.5	262.0	50.4	27.1
3084	COASTAL ANDHRA PRADESH	1903	0.8	13.3	0.2	6.2	73.4	154.0	248.6	258.0	216.5	159.1	173.9	12.1
3085	COASTAL ANDHRA PRADESH	1904	1.3	0.0	5.4	3.0	136.3	107.8	120.2	117.7	116.8	240.9	0.0	10.7
3086	COASTAL ANDHRA PRADESH	1905	1.1	16.7	68.0	37.0	68.8	84.4	64.6	210.8	170.2	66.0	7.4	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3192	COASTAL ANDHRA PRADESH	2011	0.0	17.9	0.9	62.3	67.9	86.8	196.0	215.8	129.7	74.6	4.9	5.0
3193	COASTAL ANDHRA PRADESH	2012	37.6	0.0	2.7	24.0	39.3	95.4	221.9	221.2	246.5	140.0	289.7	0.0
3194	COASTAL ANDHRA PRADESH	2013	2.0	29.6	0.2	48.0	28.2	127.5	162.4	123.1	132.0	411.5	53.1	2.8
3195	COASTAL ANDHRA PRADESH	2014	0.4	1.2	9.1	6.0	112.9	45.7	151.8	177.8	144.5	195.6	23.7	6.4
3196	COASTAL ANDHRA PRADESH	2015	2.0	0.6	5.5	32.3	34.1	283.8	116.0	192.0	201.8	59.7	81.2	2.0

115 rows × 14 columns

In [86]:
```
data3.isna().sum()
```

Out[86]:
```
SUBDIVISION    0
YEAR           0
JAN            0
FEB            0
MAR            0
APR            0
MAY            0
JUN            0
JUL            0
AUG            0
SEP            0
OCT            0
NOV            0
DEC            0
dtype: int64
```

In [87]:
```
data3['ANNUAL RAIN']=data3.apply(lambda row:row.JAN+ row.FEB+row.MAR+row.APR+row.MAY+row.JUN+row.JUL+row.AUG
```

In [88]:
```
import warnings
warnings.filterwarnings("ignore")
```

In [89]:
```
data3
```

Out[89]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL RAIN
3082	COASTAL ANDHRA PRADESH	1901	18.8	80.9	7.2	28.7	68.7	77.7	113.0	133.7	125.3	173.4	164.8	1.5	993.7
3083	COASTAL ANDHRA PRADESH	1902	2.0	0.0	2.8	23.9	37.6	72.6	144.5	236.1	204.5	262.0	50.4	27.1	1063.5
3084	COASTAL ANDHRA PRADESH	1903	0.8	13.3	0.2	6.2	73.4	154.0	248.6	258.0	216.5	159.1	173.9	12.1	1316.1
3085	COASTAL ANDHRA PRADESH	1904	1.3	0.0	5.4	3.0	136.3	107.8	120.2	117.7	116.8	240.9	0.0	10.7	860.1
3086	COASTAL ANDHRA PRADESH	1905	1.1	16.7	68.0	37.0	68.8	84.4	64.6	210.8	170.2	66.0	7.4	0.0	795.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3192	COASTAL ANDHRA PRADESH	2011	0.0	17.9	0.9	62.3	67.9	86.8	196.0	215.8	129.7	74.6	4.9	5.0	861.8
3193	COASTAL ANDHRA PRADESH	2012	37.6	0.0	2.7	24.0	39.3	95.4	221.9	221.2	246.5	140.0	289.7	0.0	1318.3
3194	COASTAL ANDHRA PRADESH	2013	2.0	29.6	0.2	48.0	28.2	127.5	162.4	123.1	132.0	411.5	53.1	2.8	1120.4
3195	COASTAL ANDHRA PRADESH	2014	0.4	1.2	9.1	6.0	112.9	45.7	151.8	177.8	144.5	195.6	23.7	6.4	875.1
3196	COASTAL ANDHRA PRADESH	2015	2.0	0.6	5.5	32.3	34.1	283.8	116.0	192.0	201.8	59.7	81.2	2.0	1011.0

115 rows × 15 columns

In [90]:
```
data4=data3.drop(['SUBDIVISION'],axis=1)
```

In [91]:
```
data4
```

Out[91]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL RAIN
3082	1901	18.8	80.9	7.2	28.7	68.7	77.7	113.0	133.7	125.3	173.4	164.8	1.5	993.7
3083	1902	2.0	0.0	2.8	23.9	37.6	72.6	144.5	236.1	204.5	262.0	50.4	27.1	1063.5
3084	1903	0.8	13.3	0.2	6.2	73.4	154.0	248.6	258.0	216.5	159.1	173.9	12.1	1316.1
3085	1904	1.3	0.0	5.4	3.0	136.3	107.8	120.2	117.7	116.8	240.9	0.0	10.7	860.1
3086	1905	1.1	16.7	68.0	37.0	68.8	84.4	64.6	210.8	170.2	66.0	7.4	0.0	795.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3192	2011	0.0	17.9	0.9	62.3	67.9	86.8	196.0	215.8	129.7	74.6	4.9	5.0	861.8
3193	2012	37.6	0.0	2.7	24.0	39.3	95.4	221.9	221.2	246.5	140.0	289.7	0.0	1318.3
3194	2013	2.0	29.6	0.2	48.0	28.2	127.5	162.4	123.1	132.0	411.5	53.1	2.8	1120.4
3195	2014	0.4	1.2	9.1	6.0	112.9	45.7	151.8	177.8	144.5	195.6	23.7	6.4	875.1
3196	2015	2.0	0.6	5.5	32.3	34.1	283.8	116.0	192.0	201.8	59.7	81.2	2.0	1011.0

115 rows × 14 columns

In [92]:
```
cor=data4.corr()
cor
```

Out[92]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	
YEAR	1.000000	0.048132	-0.073781	0.031480	-0.020721	0.049330	0.009624	0.139075	0.168372	0.019305	0.004070	-0.037462	0.0
JAN	0.048132	1.000000	-0.062433	-0.019689	-0.065377	-0.025647	-0.148502	0.016738	0.004698	-0.006220	-0.076399	0.124096	0.0
FEB	-0.073781	-0.062433	1.000000	0.173153	0.051741	0.252374	-0.002180	-0.123334	0.007815	-0.104363	0.001682	0.094418	-0.0
MAR	0.031480	-0.019689	0.173153	1.000000	0.061019	0.222856	-0.041628	-0.031694	0.078904	-0.088769	-0.135138	-0.124286	-0.0
APR	-0.020721	-0.065377	0.051741	0.061019	1.000000	-0.066308	-0.122411	-0.125468	-0.024127	-0.063572	-0.098969	-0.168154	-0.0
MAY	0.049330	-0.025647	0.252374	0.222856	-0.066308	1.000000	-0.049903	-0.092140	0.106612	-0.127919	0.132451	0.096951	0.0
JUN	0.009624	-0.148502	-0.002180	-0.041628	-0.122411	-0.049903	1.000000	0.002577	0.164741	0.009614	-0.000123	-0.059916	0.0
JUL	0.139075	0.016738	-0.123334	-0.031694	-0.125468	-0.092140	0.002577	1.000000	0.250009	0.110985	0.077776	0.027248	0.2
AUG	0.168372	0.004698	0.007815	0.078904	-0.024127	0.106612	0.164741	0.250009	1.000000	0.016081	-0.105192	-0.044359	0.0
SEP	0.019305	-0.006220	-0.104363	-0.088769	-0.063572	-0.127919	0.009614	0.110985	0.016081	1.000000	0.006472	-0.006844	-0.0
OCT	0.004070	-0.076399	0.001682	-0.135138	-0.098969	0.132451	-0.000123	0.077776	-0.105192	0.006472	1.000000	-0.005399	-0.0

In [138]:
```
data5=data4.drop(['JAN','FEB','MAR','APR','MAY'],axis=1)
```

In [139]:
```
data5
```

Out[139]:

	YEAR	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL RAIN
3082	1901	77.7	113.0	133.7	125.3	173.4	164.8	1.5	993.7
3083	1902	72.6	144.5	236.1	204.5	262.0	50.4	27.1	1063.5
3084	1903	154.0	248.6	258.0	216.5	159.1	173.9	12.1	1316.1
3085	1904	107.8	120.2	117.7	116.8	240.9	0.0	10.7	860.1
3086	1905	84.4	64.6	210.8	170.2	66.0	7.4	0.0	795.0
...	...	...	...	...	...	...	...	...	...
3192	2011	86.8	196.0	215.8	129.7	74.6	4.9	5.0	861.8
3193	2012	95.4	221.9	221.2	246.5	140.0	289.7	0.0	1318.3
3194	2013	127.5	162.4	123.1	132.0	411.5	53.1	2.8	1120.4
3195	2014	45.7	151.8	177.8	144.5	195.6	23.7	6.4	875.1
3196	2015	283.8	116.0	192.0	201.8	59.7	81.2	2.0	1011.0

115 rows × 9 columns

In [140]:
```
data5['SWM']=data5.apply(lambda row:row.JUN+row.JUL+row.AUG+row.SEP,axis=1)
```

In [141]:
```
data5
```

Out[141]:

	YEAR	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL RAIN	SWM
3082	1901	77.7	113.0	133.7	125.3	173.4	164.8	1.5	993.7	449.7
3083	1902	72.6	144.5	236.1	204.5	262.0	50.4	27.1	1063.5	657.7
3084	1903	154.0	248.6	258.0	216.5	159.1	173.9	12.1	1316.1	877.1
3085	1904	107.8	120.2	117.7	116.8	240.9	0.0	10.7	860.1	462.5
3086	1905	84.4	64.6	210.8	170.2	66.0	7.4	0.0	795.0	530.0
...	...	...	...	...	...	...	...	...	...	...
3192	2011	86.8	196.0	215.8	129.7	74.6	4.9	5.0	861.8	628.3
3193	2012	95.4	221.9	221.2	246.5	140.0	289.7	0.0	1318.3	785.0
3194	2013	127.5	162.4	123.1	132.0	411.5	53.1	2.8	1120.4	545.0
3195	2014	45.7	151.8	177.8	144.5	195.6	23.7	6.4	875.1	519.8
3196	2015	283.8	116.0	192.0	201.8	59.7	81.2	2.0	1011.0	793.6

115 rows × 10 columns

In [144]:
```
data5['NEM']=data5.apply(lambda row:row.DEC+row.OCT+row.NOV,axis=1)
```

In [145]:
```
data5
```

Out[145]:

	YEAR	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL RAIN	SWM	NEM
3082	1901	77.7	113.0	133.7	125.3	173.4	164.8	1.5	993.7	449.7	339.7
3083	1902	72.6	144.5	236.1	204.5	262.0	50.4	27.1	1063.5	657.7	339.5
3084	1903	154.0	248.6	258.0	216.5	159.1	173.9	12.1	1316.1	877.1	345.1
3085	1904	107.8	120.2	117.7	116.8	240.9	0.0	10.7	860.1	462.5	251.6
3086	1905	84.4	64.6	210.8	170.2	66.0	7.4	0.0	795.0	530.0	73.4
...	...	...	...	...	...	...	...	...	...	...	...
3192	2011	86.8	196.0	215.8	129.7	74.6	4.9	5.0	861.8	628.3	84.5
3193	2012	95.4	221.9	221.2	246.5	140.0	289.7	0.0	1318.3	785.0	429.7
3194	2013	127.5	162.4	123.1	132.0	411.5	53.1	2.8	1120.4	545.0	467.4
3195	2014	45.7	151.8	177.8	144.5	195.6	23.7	6.4	875.1	519.8	225.7
3196	2015	283.8	116.0	192.0	201.8	59.7	81.2	2.0	1011.0	793.6	142.9

115 rows × 11 columns

In [146]:
```
data6=data5.drop(['JUN','JUL','AUG','SEP','OCT','NOV','DEC'],axis=1)
```

In [147]:
```
cor=data6.corr()
cor
```

Out[147]:

	YEAR	ANNUAL RAIN	SWM	NEM
YEAR	1.000000	0.102054	0.146870	-0.011896
ANNUAL RAIN	0.102054	1.000000	0.596706	0.677796
SWM	0.146870	0.596706	1.000000	-0.007822
NEM	-0.011896	0.677796	-0.007822	1.000000

In [148]:
```
import seaborn as sns
sns.heatmap(cor,vmax=1,vmin=-1,annot=True,linewidth=5,cmap='bwr')
```

Out[148]: <Axes: >