importing pandas

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

reading csv data

In [2]:
```
data=pd.read_csv("/home/placement/Downloads/arunachal.csv")
```

In [3]:
```
data.describe()
```

Out[3]:

	Unnamed: 0	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG
count	91.00000	91.000000	90.000000	90.000000	89.000000	91.000000	91.000000	90.000000	90.000000	91.000000
mean	155.00000	1962.747253	48.598889	93.966667	154.446067	262.990110	364.651648	659.556667	711.963333	502.163736
std	26.41338	27.695003	34.687078	46.258375	87.918484	113.395773	181.095447	311.642230	356.372598	275.716730
min	110.00000	1916.000000	1.800000	6.100000	28.500000	94.700000	101.800000	239.400000	233.000000	172.400000
25%	132.50000	1938.500000	20.075000	65.625000	101.700000	180.600000	237.150000	425.675000	442.150000	301.100000
50%	155.00000	1964.000000	45.400000	87.600000	141.700000	245.400000	314.600000	545.750000	613.000000	411.600000
75%	177.50000	1986.500000	65.150000	120.400000	189.600000	335.300000	447.050000	840.400000	922.075000	669.200000
max	200.00000	2009.000000	164.500000	208.500000	605.600000	595.100000	1168.600000	1609.900000	2362.800000	1664.600000

In [4]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 91 entries, 0 to 90
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   Unnamed: 0   91 non-null     int64  
 1   SUBDIVISION  91 non-null     object 
 2   YEAR         91 non-null     int64  
 3   JAN          90 non-null     float64
 4   FEB          90 non-null     float64
 5   MAR          89 non-null     float64
 6   APR          91 non-null     float64
 7   MAY          91 non-null     float64
 8   JUN          90 non-null     float64
 9   JUL          90 non-null     float64
 10  AUG          91 non-null     float64
 11  SEP          91 non-null     float64
 12  OCT          89 non-null     float64
 13  NOV          89 non-null     float64
 14  DEC          89 non-null     float64
 15  ANNUAL       85 non-null     float64
 16  Jan-Feb      90 non-null     float64
 17  Mar-May      89 non-null     float64
 18  Jun-Sep      89 non-null     float64
 19  Oct-Dec      88 non-null     float64
dtypes: float64(17), int64(2), object(1)
memory usage: 14.3+ KB
```

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

Out[5]:

	Unnamed: 0	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb
86	196	ARUNACHAL PRADESH	2005	48.4	167.6	229.5	195.3	179.8	269.3	430.8	400.0	243.6	139.3	28.6	3.3	2335.5	216.0
87	197	ARUNACHAL PRADESH	2006	6.0	103.7	63.3	202.7	321.7	520.4	382.2	227.6	263.2	77.2	69.7	21.7	2259.6	109.7
88	198	ARUNACHAL PRADESH	2007	13.4	97.4	48.1	292.4	250.4	530.2	761.0	364.6	529.3	102.6	24.3	6.9	3020.7	110.8
89	199	ARUNACHAL PRADESH	2008	76.7	39.7	122.6	192.4	185.0	423.6	456.1	439.3	189.7	115.1	1.7	2.6	2244.4	116.4
90	200	ARUNACHAL PRADESH	2009	18.0	92.8	72.1	132.7	189.9	259.1	329.9	370.3	152.5	82.9	33.9	15.9	1749.9	110.8

Listing the data

In [6]:
```
list(data.columns)
```

Out[6]:
```
['Unnamed: 0',
 'SUBDIVISION',
 'YEAR',
 'JAN',
 'FEB',
 'MAR',
 'APR',
 'MAY',
 'JUN',
 'JUL',
 'AUG',
 'SEP',
 'OCT',
 'NOV',
 'DEC',
 'ANNUAL',
 'Jan-Feb',
 'Mar-May',
 'Jun-Sep',
 'Oct-Dec']
```

Dropping umnamed column

In [7]:
```
drop=data.drop(["Unnamed: 0"],axis=1)
```

In [8]:
```
c=drop
c
```

Out[8]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	
0	ARUNACHAL PRADESH	1916	48.1	69.8	71.1	316.1	424.6	1124.9	NaN	629.7	333.9	NaN	NaN	NaN	NaN	117.9	811.8	
1	ARUNACHAL PRADESH	1917	21.4	164.5	NaN	269.6	107.9	823.8	909.1	628.4	411.5	199.3	63.5	0.0	NaN	185.9	NaN	2
2	ARUNACHAL PRADESH	1918	10.4	11.0	191.2	144.6	861.1	1609.9	1303.0	692.6	515.8	125.2	7.8	13.7	5486.3	21.4	1196.9	4
3	ARUNACHAL PRADESH	1919	34.5	67.8	28.5	256.9	420.6	973.6	999.0	286.7	628.7	948.3	40.7	8.6	4693.9	102.3	706.0	2
4	ARUNACHAL PRADESH	1920	14.0	196.3	605.6	364.7	173.6	840.6	535.4	896.5	376.7	103.3	0.0	0.0	4106.7	210.3	1143.9	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
86	ARUNACHAL PRADESH	2005	48.4	167.6	229.5	195.3	179.8	269.3	430.8	400.0	243.6	139.3	28.6	3.3	2335.5	216.0	604.6	1
87	ARUNACHAL PRADESH	2006	6.0	103.7	63.3	202.7	321.7	520.4	382.2	227.6	263.2	77.2	69.7	21.7	2259.6	109.7	587.7	1
88	ARUNACHAL PRADESH	2007	13.4	97.4	48.1	292.4	250.4	530.2	761.0	364.6	529.3	102.6	24.3	6.9	3020.7	110.8	590.9	2
89	ARUNACHAL PRADESH	2008	76.7	39.7	122.6	192.4	185.0	423.6	456.1	439.3	189.7	115.1	1.7	2.6	2244.4	116.4	499.9	1
90	ARUNACHAL PRADESH	2009	18.0	92.8	72.1	132.7	189.9	259.1	329.9	370.3	152.5	82.9	33.9	15.9	1749.9	110.8	394.7	1

91 rows × 19 columns

checking the nulls

In [9]:
```
c.isna().sum()
```

Out[9]:
```
SUBDIVISION    0
YEAR           0
JAN            1
FEB            1
MAR            2
APR            0
MAY            0
JUN            1
JUL            1
AUG            0
SEP            0
OCT            2
NOV            2
DEC            2
ANNUAL         6
Jan-Feb        1
Mar-May        2
Jun-Sep        2
Oct-Dec        3
dtype: int64
```

In [10]:
```
c=pd.get_dummies(c)
c
```

Out[10]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec	S
0	1916	48.1	69.8	71.1	316.1	424.6	1124.9	NaN	629.7	333.9	NaN	NaN	NaN	NaN	117.9	811.8	NaN	NaN	
1	1917	21.4	164.5	NaN	269.6	107.9	823.8	909.1	628.4	411.5	199.3	63.5	0.0	NaN	185.9	NaN	2772.8	262.8	
2	1918	10.4	11.0	191.2	144.6	861.1	1609.9	1303.0	692.6	515.8	125.2	7.8	13.7	5486.3	21.4	1196.9	4121.3	146.7	
3	1919	34.5	67.8	28.5	256.9	420.6	973.6	999.0	286.7	628.7	948.3	40.7	8.6	4693.9	102.3	706.0	2888.0	997.6	
4	1920	14.0	196.3	605.6	364.7	173.6	840.6	535.4	896.5	376.7	103.3	0.0	0.0	4106.7	210.3	1143.9	2649.2	103.3	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
86	2005	48.4	167.6	229.5	195.3	179.8	269.3	430.8	400.0	243.6	139.3	28.6	3.3	2335.5	216.0	604.6	1343.7	171.2	
87	2006	6.0	103.7	63.3	202.7	321.7	520.4	382.2	227.6	263.2	77.2	69.7	21.7	2259.6	109.7	587.7	1393.5	168.7	
88	2007	13.4	97.4	48.1	292.4	250.4	530.2	761.0	364.6	529.3	102.6	24.3	6.9	3020.7	110.8	590.9	2185.1	133.9	
89	2008	76.7	39.7	122.6	192.4	185.0	423.6	456.1	439.3	189.7	115.1	1.7	2.6	2244.4	116.4	499.9	1508.7	119.4	
90	2009	18.0	92.8	72.1	132.7	189.9	259.1	329.9	370.3	152.5	82.9	33.9	15.9	1749.9	110.8	394.7	1111.8	132.7	

91 rows × 19 columns

In [11]:
```
c=c.fillna(c.mean())
c
```

Out[11]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Ja F
0	1916	48.1	69.8	71.100000	316.1	424.6	1124.9	711.963333	629.7	333.9	200.37191	36.257303	24.91573	3475.443529	117
1	1917	21.4	164.5	154.446067	269.6	107.9	823.8	909.100000	628.4	411.5	199.30000	63.500000	0.00000	3475.443529	185
2	1918	10.4	11.0	191.200000	144.6	861.1	1609.9	1303.000000	692.6	515.8	125.20000	7.800000	13.70000	5486.300000	21
3	1919	34.5	67.8	28.500000	256.9	420.6	973.6	999.000000	286.7	628.7	948.30000	40.700000	8.60000	4693.900000	102
4	1920	14.0	196.3	605.600000	364.7	173.6	840.6	535.400000	896.5	376.7	103.30000	0.000000	0.00000	4106.700000	210
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
86	2005	48.4	167.6	229.500000	195.3	179.8	269.3	430.800000	400.0	243.6	139.30000	28.600000	3.30000	2335.500000	216
87	2006	6.0	103.7	63.300000	202.7	321.7	520.4	382.200000	227.6	263.2	77.20000	69.700000	21.70000	2259.600000	109
88	2007	13.4	97.4	48.100000	292.4	250.4	530.2	761.000000	364.6	529.3	102.60000	24.300000	6.90000	3020.700000	110
89	2008	76.7	39.7	122.600000	192.4	185.0	423.6	456.100000	439.3	189.7	115.10000	1.700000	2.60000	2244.400000	116
90	2009	18.0	92.8	72.100000	132.7	189.9	259.1	329.900000	370.3	152.5	82.90000	33.900000	15.90000	1749.900000	110

91 rows × 19 columns

Importing warnings

In [12]:
```
import warnings
```

In [13]:
```
warnings.filterwarnings("ignore")
#ignores the warnings
```

In [14]:
```
c
```

Out[14]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Ja F
0	1916	48.1	69.8	71.100000	316.1	424.6	1124.9	711.963333	629.7	333.9	200.37191	36.257303	24.91573	3475.443529	117
1	1917	21.4	164.5	154.446067	269.6	107.9	823.8	909.100000	628.4	411.5	199.30000	63.500000	0.00000	3475.443529	185
2	1918	10.4	11.0	191.200000	144.6	861.1	1609.9	1303.000000	692.6	515.8	125.20000	7.800000	13.70000	5486.300000	21
3	1919	34.5	67.8	28.500000	256.9	420.6	973.6	999.000000	286.7	628.7	948.30000	40.700000	8.60000	4693.900000	102
4	1920	14.0	196.3	605.600000	364.7	173.6	840.6	535.400000	896.5	376.7	103.30000	0.000000	0.00000	4106.700000	210
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
86	2005	48.4	167.6	229.500000	195.3	179.8	269.3	430.800000	400.0	243.6	139.30000	28.600000	3.30000	2335.500000	216
87	2006	6.0	103.7	63.300000	202.7	321.7	520.4	382.200000	227.6	263.2	77.20000	69.700000	21.70000	2259.600000	109
88	2007	13.4	97.4	48.100000	292.4	250.4	530.2	761.000000	364.6	529.3	102.60000	24.300000	6.90000	3020.700000	110
89	2008	76.7	39.7	122.600000	192.4	185.0	423.6	456.100000	439.3	189.7	115.10000	1.700000	2.60000	2244.400000	116
90	2009	18.0	92.8	72.100000	132.7	189.9	259.1	329.900000	370.3	152.5	82.90000	33.900000	15.90000	1749.900000	110

91 rows × 19 columns

correlation

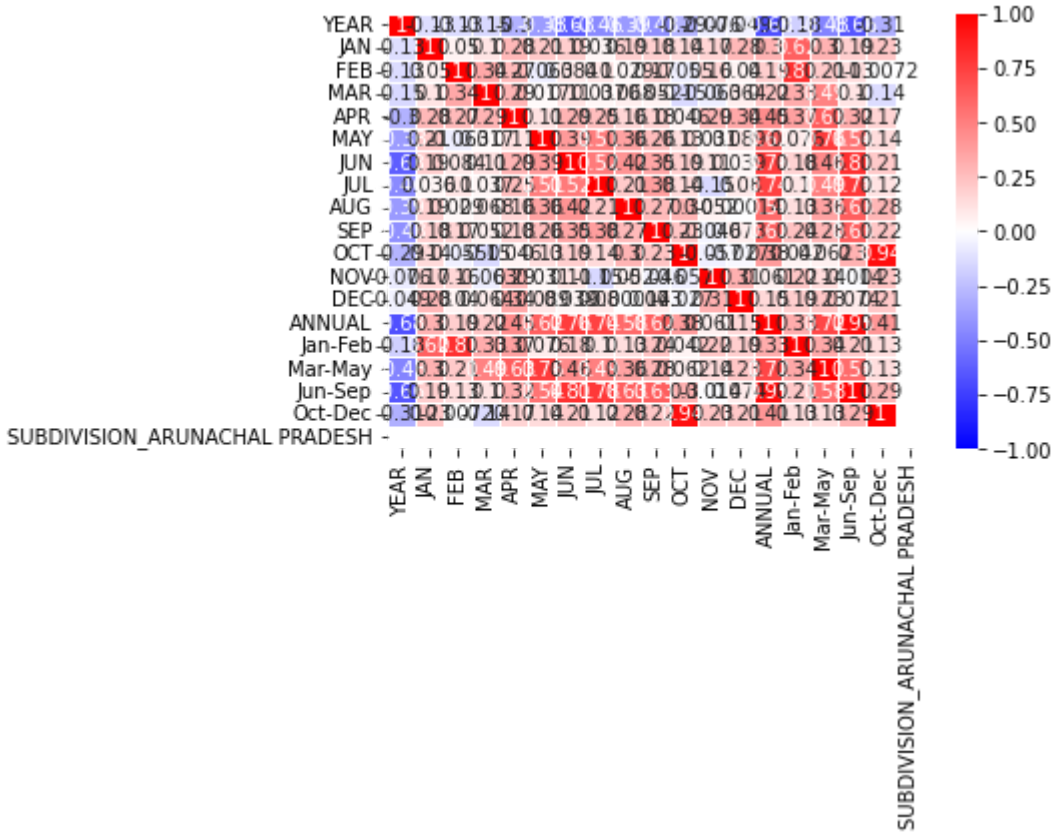
In [15]:
```
cor_mat=c.corr()
cor_mat
```

Out[15]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
YEAR	1.000000	-0.129747	-0.134367	-0.151211	-0.301073	-0.384602	-0.629752	-0.458136	-0.394444	-0.431541
JAN	-0.129747	1.000000	0.049703	0.102293	0.275434	0.213184	0.187787	0.035809	0.186374	0.180082
FEB	-0.134367	0.049703	1.000000	0.341841	0.268473	-0.063203	0.084120	0.101055	0.028858	0.168443
MAR	-0.151211	0.102293	0.341841	1.000000	0.292034	0.016967	0.109636	0.037348	0.068452	0.052290
APR	-0.301073	0.275434	0.268473	0.292034	1.000000	0.114128	0.289865	0.251932	0.157620	0.176335
MAY	-0.384602	0.213184	-0.063203	0.016967	0.114128	1.000000	0.393266	0.506670	0.363992	0.258744
JUN	-0.629752	0.187787	0.084120	0.109636	0.289865	0.393266	1.000000	0.521139	0.415815	0.346802
JUL	-0.458136	0.035809	0.101055	0.037348	0.251932	0.506670	0.521139	1.000000	0.210298	0.380633
AUG	-0.394444	0.186374	0.028858	0.068452	0.157620	0.363992	0.415815	0.210298	1.000000	0.269123
SEP	-0.431541	0.180082	0.168443	0.052290	0.176335	0.258744	0.346802	0.380633	0.269123	1.000000
OCT	-0.289344	0.144590	-0.054795	-0.148231	0.045969	0.127820	0.192367	0.144446	0.296349	0.227094
NOV	-0.076280	0.165581	0.160783	-0.063310	0.288599	0.031172	0.109367	-0.151307	0.052211	-0.046211
DEC	-0.048730	0.277939	0.040145	0.064440	0.337215	0.089220	0.038968	0.079788	-0.000140	0.072701
ANNUAL	-0.682590	0.301475	0.192391	0.219965	0.446343	0.621016	0.762945	0.739870	0.582262	0.622284
Jan-Feb	-0.181185	0.624899	0.810800	0.327051	0.371221	0.075552	0.175902	0.100006	0.131836	0.237243
Mar-May	-0.482071	0.303795	0.210505	0.488820	0.634120	0.759110	0.457520	0.492636	0.360978	0.281265
Jun-Sep	-0.662652	0.189225	0.129102	0.104851	0.323289	0.541487	0.810782	0.776939	0.632006	0.625144
Oct-Dec	-0.305427	0.226541	-0.007182	-0.139970	0.168056	0.139581	0.210311	0.115344	0.282621	0.224471
SUBDIVISION_ARUNACHAL PRADESH	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [24]:
```
import seaborn as sns
sns.heatmap(cor_mat,vmax=1,vmin=-1,annot=True,linewidth=1,cmap='bwr')#plotting of graph using seab
```

Out[24]: <Axes: >



In [17]:
```
data.isna().sum()
```

Out[17]:
```
Unnamed: 0     0
SUBDIVISION    0
YEAR           0
JAN            1
FEB            1
MAR            2
APR            0
MAY            0
JUN            1
JUL            1
AUG            0
SEP            0
OCT            2
NOV            2
DEC            2
ANNUAL         6
Jan-Feb        1
Mar-May        2
Jun-Sep        2
Oct-Dec        3
dtype: int64
```

In [ ]:

In [ ]: