

In [1]:

```
import pandas as pd
```

In [2]:

```
data=pd.read_csv("/home/placement/Downloads/arunachal.csv")
```

In [6]:

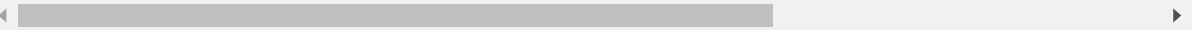
```
data1=pd.read_csv("/home/placement/Downloads/rainfall in india 1901-2015.csv")
```

In [4]:

```
data.head()
```

Out[4]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	N
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	55
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	35
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	28
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	30
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	2



In [7]:

```
data1.head()
```

Out[7]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	N
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	55
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	35
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	28
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	30
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	2

In [12]:

```
list(data)
```

Out[12]:

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

In [13]:

```
list(data1)
```

Out[13]:

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

In [14]:

```
data1=data.loc[(data.YEAR < 2010)] #data till year 2010, can change here
```

In [15]:

```
data1.tail()
```

Out[15]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
4105	LAKSHADWEEP	2005	17.6	11.1	0.0	37.0	92.8	248.5	378.9	102.4	278.0	164.2
4106	LAKSHADWEEP	2006	20.1	0.0	33.0	0.3	327.9	286.9	172.3	150.7	318.5	119.1
4107	LAKSHADWEEP	2007	2.5	4.2	0.2	22.2	166.2	573.4	427.4	294.7	457.5	256.1
4108	LAKSHADWEEP	2008	5.5	19.8	120.7	15.8	180.4	254.6	363.9	206.6	108.9	252.9
4109	LAKSHADWEEP	2009	4.7	1.5	0.1	18.1	162.1	401.2	266.4	185.0	145.1	87.4

In [18]:

```
data1['SUBDIVISION'].unique()
```

Out[18]:

```
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 & CHANDIGAR  
H',  
      'PUNJAB', 'HIMACHAL PRADESH', 'JAMMU & KASHMIR', 'WEST RAJASTHA  
N',  
      '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 NAD  
U',  
      'COASTAL KARNATAKA', 'NORTH INTERIOR KARNATAKA',  
      'SOUTH INTERIOR KARNATAKA', 'KERALA', 'LAKSHADWEEP'], dtype=obj  
ect)
```

In [25]:

```
data2=data1.loc[(data1.SUBDIVISION == 'ARUNACHAL PRADESH')]
```

In [26]:

data2

Out[26]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
110	ARUNACHAL PRADESH	1916	48.1	69.8	71.1	316.1	424.6	1124.9	NaN	629.7	333.9	NaN
111	ARUNACHAL PRADESH	1917	21.4	164.5	NaN	269.6	107.9	823.8	909.1	628.4	411.5	199.1
112	ARUNACHAL PRADESH	1918	10.4	11.0	191.2	144.6	861.1	1609.9	1303.0	692.6	515.8	125.1
113	ARUNACHAL PRADESH	1919	34.5	67.8	28.5	256.9	420.6	973.6	999.0	286.7	628.7	948.1
114	ARUNACHAL PRADESH	1920	14.0	196.3	605.6	364.7	173.6	840.6	535.4	896.5	376.7	103.1
...
196	ARUNACHAL PRADESH	2005	48.4	167.6	229.5	195.3	179.8	269.3	430.8	400.0	243.6	139.1
197	ARUNACHAL PRADESH	2006	6.0	103.7	63.3	202.7	321.7	520.4	382.2	227.6	263.2	77.1
198	ARUNACHAL PRADESH	2007	13.4	97.4	48.1	292.4	250.4	530.2	761.0	364.6	529.3	102.6
199	ARUNACHAL PRADESH	2008	76.7	39.7	122.6	192.4	185.0	423.6	456.1	439.3	189.7	115.1
200	ARUNACHAL PRADESH	2009	18.0	92.8	72.1	132.7	189.9	259.1	329.9	370.3	152.5	82.9

91 rows × 19 columns



In [27]:

```
data2.isna().sum()
```

Out[27]:

```
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 [28]:

```
data3=data.loc[(data.SUBDIVISION == 'ARUNACHAL PRADESH')]
```

In [30]:

```
data3.isna().sum()
```

Out[30]:

```
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 [32]:

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

In [33]:

```
data4
```

Out[33]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
110	1916	48.1	69.8	71.1	316.1	424.6	1124.9	NaN	629.7	333.9	NaN	NaN	NaN
111	1917	21.4	164.5	NaN	269.6	107.9	823.8	909.1	628.4	411.5	199.3	63.5	0.0
112	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
113	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
114	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
...
202	2011	40.0	51.3	174.5	240.8	219.6	288.4	531.4	277.6	286.7	51.9	16.2	15.2
203	2012	57.8	35.8	134.2	403.4	187.4	645.8	638.9	316.0	724.9	248.1	22.0	26.2
204	2013	18.5	40.5	115.1	175.1	335.8	290.0	329.6	230.2	316.1	164.1	13.3	14.6
205	2014	19.0	101.9	80.3	86.7	299.0	415.8	392.4	599.6	343.0	35.1	20.1	10.2
206	2015	30.8	47.5	97.5	287.1	238.9	637.9	329.3	595.5	374.2	65.2	33.8	29.8

97 rows × 18 columns



In [36]:

```
cor=data.corr()  
cor
```

/tmp/ipykernel_7018/4173678507.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
cor=data.corr()

Out[36]:

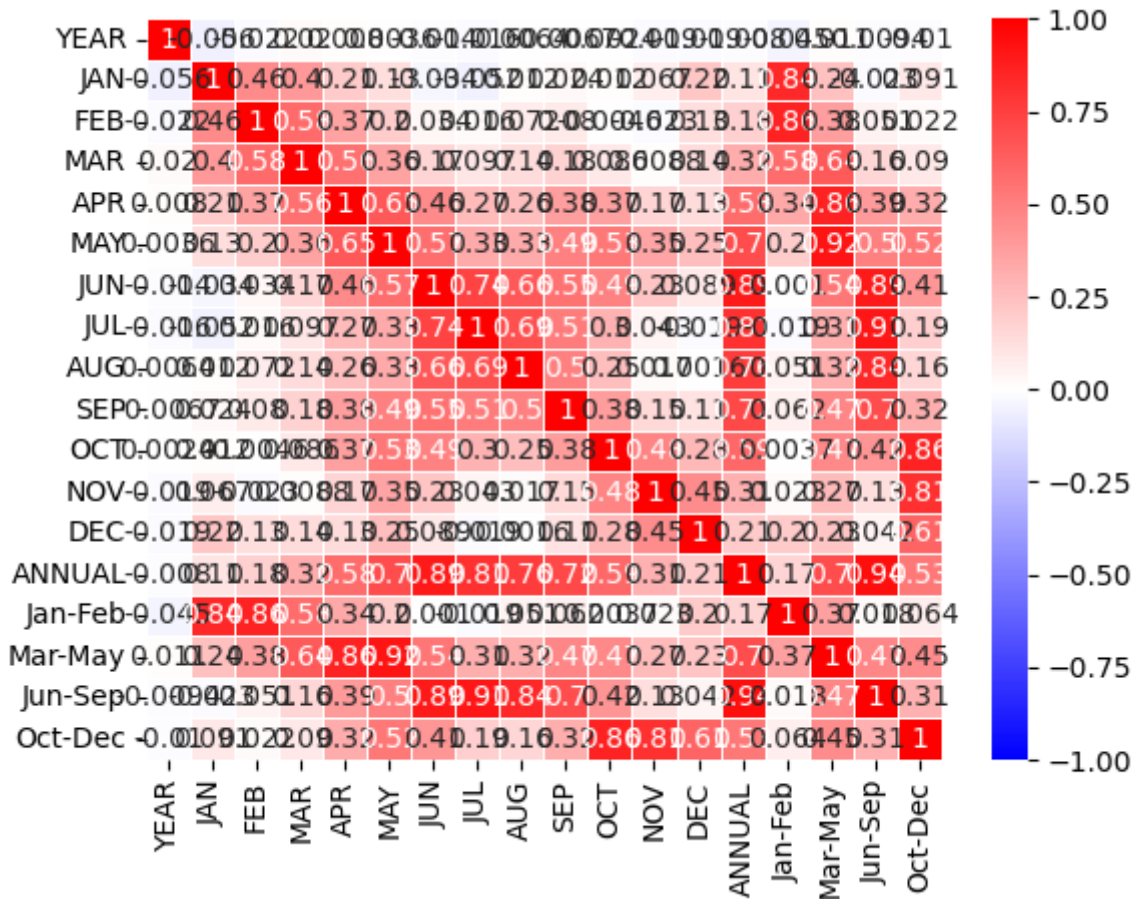
	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL
YEAR	1.000000	-0.056235	-0.022144	0.020338	0.008007	0.003594	-0.013594	-0.016240
JAN	-0.056235	1.000000	0.456183	0.398502	0.209302	0.129622	-0.033725	-0.051642
FEB	-0.022144	0.456183	1.000000	0.579819	0.367114	0.203062	0.033703	0.016235
MAR	0.020338	0.398502	0.579819	1.000000	0.556856	0.362815	0.165857	0.097334
APR	0.008007	0.209302	0.367114	0.556856	1.000000	0.650595	0.457091	0.268097
MAY	0.003594	0.129622	0.203062	0.362815	0.650595	1.000000	0.567618	0.332283
JUN	-0.013594	-0.033725	0.033703	0.165857	0.457091	0.567618	1.000000	0.741285
JUL	-0.016240	-0.051642	0.016235	0.097334	0.268097	0.332283	0.741285	1.000000
AUG	0.006442	0.011952	0.072159	0.135071	0.256168	0.329499	0.655142	0.686662
SEP	-0.006670	0.024289	0.080148	0.178904	0.382525	0.492378	0.551890	0.513067
OCT	0.002406	0.012374	-0.004581	0.086187	0.368886	0.529342	0.490393	0.299221
NOV	-0.018776	0.067281	-0.023413	0.008814	0.165642	0.351931	0.229718	0.042671
DEC	-0.019139	0.219701	0.132570	0.136328	0.132892	0.250112	0.088782	-0.019427
ANNUAL	-0.008044	0.105696	0.181563	0.322199	0.577573	0.698013	0.891303	0.812279
Jan-Feb	-0.044653	0.842390	0.863815	0.576366	0.340841	0.196168	0.001016	-0.019157
Mar-May	0.010637	0.242256	0.382620	0.642294	0.864172	0.915019	0.538562	0.313726
Jun-Sep	-0.009418	-0.022748	0.051066	0.162055	0.394859	0.496164	0.893968	0.907723
Oct-Dec	-0.010155	0.090932	0.021878	0.090108	0.321407	0.523684	0.409050	0.190400

In [39]:

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

Out[39]:

<Axes: >



In []:

