

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

```
Out[2]:
```

	Unnamed: 0	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May
0	110	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	111	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	112	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
3	113	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
4	114	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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
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	604.6
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	587.7
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	590.9
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	499.9
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	394.7

91 rows × 20 columns



## preprocessing the data

```
In [3]: data.shape
```

```
Out[3]: (91, 20)
```

```
In [4]: data.columns
```

```
Out[4]: Index(['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'],  
             dtype='object')
```

In [5]: 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 [6]: data.describe()

Out[6]:

	Unnamed: 0	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
count	91.00000	91.000000	90.000000	90.000000	89.000000	91.000000	91.000000	90.000000	90.000000	91.000000	91.000000
mean	155.00000	1962.747253	48.598889	93.966667	154.446067	262.990110	364.651648	659.556667	711.963333	502.163736	433.273626
std	26.41338	27.695003	34.687078	46.258375	87.918484	113.395773	181.095447	311.642230	356.372598	275.716730	204.991358
min	110.00000	1916.000000	1.800000	6.100000	28.500000	94.700000	101.800000	239.400000	233.000000	172.400000	152.500000
25%	132.50000	1938.500000	20.075000	65.625000	101.700000	180.600000	237.150000	425.675000	442.150000	301.100000	282.150000
50%	155.00000	1964.000000	45.400000	87.600000	141.700000	245.400000	314.600000	545.750000	613.000000	411.600000	384.300000
75%	177.50000	1986.500000	65.150000	120.400000	189.600000	335.300000	447.050000	840.400000	922.075000	669.200000	521.150000
max	200.00000	2009.000000	164.500000	208.500000	605.600000	595.100000	1168.600000	1609.900000	2362.800000	1664.600000	1222.000000



## drop()

In [7]: data1=data.drop(columns='Unnamed: 0')

In [8]: data1

Out[8]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec
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	NaN	NaN
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	2772.8	262.
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	4121.3	146.
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	2888.0	997.
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	2649.2	103.
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
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	1343.7	171.
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	1393.5	168.
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	2185.1	133.
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	1508.7	119.
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	1111.8	132.

91 rows × 19 columns



map()

```
In [9]: data['SUBDIVISION']=data['SUBDIVISION'].map({'ARUNACHAL PRADESH':1})
```

```
In [10]: data
```

```
Out[10]:
```

	Unnamed: 0	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May
0	110	1	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	111	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
2	112	1	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
3	113	1	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
4	114	1	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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
86	196	1	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
87	197	1	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
88	198	1	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
89	199	1	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
90	200	1	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

91 rows × 20 columns



isna()

```
In [11]: data1.isna().sum()
```

```
Out[11]: 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
```

## fillna()

```
In [12]: data2=data1.fillna(data.mean())
```

```
In [13]: data2.isna().sum()
```

```
Out[13]: 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  
        ANNUAL         0  
        Jan-Feb        0  
        Mar-May        0  
        Jun-Sep        0  
        Oct-Dec        0  
        dtype: int64
```



In [14]: data2

Out[14]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
0	ARUNACHAL PRADESH	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
1	ARUNACHAL PRADESH	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
2	ARUNACHAL PRADESH	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
3	ARUNACHAL PRADESH	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
4	ARUNACHAL PRADESH	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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
86	ARUNACHAL PRADESH	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
87	ARUNACHAL PRADESH	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
88	ARUNACHAL PRADESH	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
89	ARUNACHAL PRADESH	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
90	ARUNACHAL PRADESH	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

91 rows × 19 columns



## Drop the unwanted columns

```
In [15]: data3=data2.drop(columns=["Jan-Feb", "Mar-May", "Jun-Sep", "Oct-Dec"])
```

```
In [16]: data3
```

```
Out[16]:
```

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
0	ARUNACHAL PRADESH	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
1	ARUNACHAL PRADESH	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
2	ARUNACHAL PRADESH	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
3	ARUNACHAL PRADESH	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
4	ARUNACHAL PRADESH	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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
86	ARUNACHAL PRADESH	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
87	ARUNACHAL PRADESH	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
88	ARUNACHAL PRADESH	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
89	ARUNACHAL PRADESH	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
90	ARUNACHAL PRADESH	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

91 rows × 15 columns

```
In [17]: data3.isna().sum()
```

```
Out[17]: 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  
        ANNUAL          0  
        dtype: int64
```

```
In [18]: data4=data3.drop(columns='SUBDIVISION')
```

In [19]: data4

Out[19]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
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
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
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
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
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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
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
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
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
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
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

91 rows × 14 columns

## Adding new columns

In [20]: data4['SWM']=data4['JUN']+data4['JUL']+data4['AUG']+data4['SEP']  
 data4['NEM']=data4['OCT']+data4['NOV']+data4['DEC']

In [21]: data4

Out[21]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	SWM
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	2800.463333
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	2772.800000
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	4121.300000
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	2888.000000
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	2649.200000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
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	1343.700000
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	1393.400000
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	2185.100000
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	1508.700000
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	1111.800000

91 rows × 16 columns



## Again drop the unwanted columns

In [22]: data5=data4.drop(columns=['JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL', 'AUG', 'SEP', 'OCT', 'NOV', 'DEC'])

In [23]: data5

Out[23]:

	YEAR	ANNUAL	SWM	NEM
0	1916	3475.443529	2800.463333	261.544944
1	1917	3475.443529	2772.800000	262.800000
2	1918	5486.300000	4121.300000	146.700000
3	1919	4693.900000	2888.000000	997.600000
4	1920	4106.700000	2649.200000	103.300000
...	...	...	...	...
86	2005	2335.500000	1343.700000	171.200000
87	2006	2259.600000	1393.400000	168.600000
88	2007	3020.700000	2185.100000	133.800000
89	2008	2244.400000	1508.700000	119.400000
90	2009	1749.900000	1111.800000	132.700000

91 rows × 4 columns

## correlation

In [24]: cor=data5.corr()

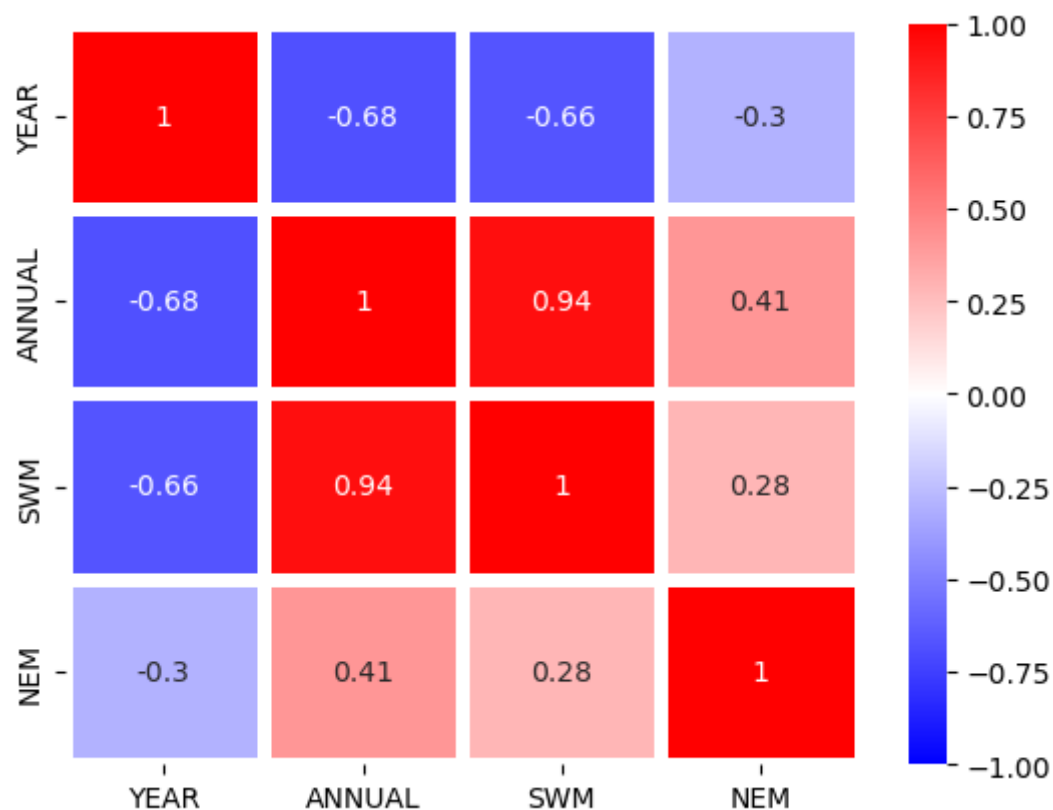
```
In [25]: cor
```

```
Out[25]:
```

	YEAR	ANNUAL	SWM	NEM
YEAR	1.000000	-0.682590	-0.664900	-0.304589
ANNUAL	-0.682590	1.000000	0.942978	0.408974
SWM	-0.664900	0.942978	1.000000	0.282742
NEM	-0.304589	0.408974	0.282742	1.000000

```
In [26]: import seaborn as sns  
sns.heatmap(cor, vmax=1, vmin=-1, annot=True, linewidth=5, cmap='bwr')
```

```
Out[26]: <Axes: >
```



In [ ]: