

# Rainfall Prediction in India

```
[163.] # Importing required packages:
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

In [164.] # Invoke data from CSV file:
rainfall_dataset=pd.read_csv('rainfall_india.csv')
print(rainfall_dataset)

JAN    FEB    MAR    APR    MAY    JUN    JUL    AUG    SEP    OCT    NOV    DEC    ANNUAL    Jan-Feb    Mar-May    Jun-Sep    Oct-Dec
0  107.3  57.9  65.2  117.0  358.5  295.5  285.0  315.2  250.9  280.5  165.2  540.7  1207.2  892.1  455.6  326.0  271.9  354.8  326.0
1  43.7  26.0  18.6  90.5  374.4  457.2  421.3  423.1  455.6  301.2  275.8  128.3  3015.7  69.7  483.5  1757.2  465.4  480.9  454.8  286.1
2  32.7  15.9  8.6  53.4  343.6  503.3  465.4  460.9  450.0  291.3  48.6  405.6  1884.4  574.7  405.6  1884.4  574.7  405.6  1884.4  574.7
3  42.2  80.8  17.4  39.5  385.4  447.0  660.1  427.8  313.6  167.1  341.1  29.8  3043.8  123.0  841.3  1848.5  231.0  666.1  427.8  313.6  167.1
4  33.3  79.5  105.9  216.5  323.0  738.3  990.9  711.2  568.0  206.9  29.5  31.7  4334.7  112.8  645.4  3008.4  268.1  454.8  326.0  271.9  354.8  326.0
636  13.4  22.1  43.6  150.4  232.6  651.6  788.9  527.3  388.4  343.2  637  2.3  1.0  1.4  46.9  227.6  959.9  1208.5  636.3  263.1  234.9
638  19.8  45.2  73.9  181.9  284.7  556.8  539.9  352.7  266.2  359.4  639  4.8  8.3  17.5  83.3  174.6  698.1  1210.4  592.9  238.7  137.1
640  20.8  54.7  11.8  48.9  371.7  238.2  287.7  217.5  163.1  137.1

NOV    DEC    ANNUAL    Jan-Feb    Mar-May    Jun-Sep    Oct-Dec
0  315.2  259.9  2805.2  165.2  540.7  1207.2  892.1
1  275.8  128.3  3015.7  69.7  483.5  1757.2  465.4
2  195.6  380.0  2313.3  48.6  405.6  1884.4  574.7
3  34.1  29.8  3043.8  123.0  841.3  1848.5  231.0
4  29.5  31.7  4334.7  112.8  645.4  3008.4  268.1
... ..
636  172.9  48.1  3302.5  35.5  426.6  2276.2  964.7
637  84.6  38.4  3021.6  3  272.9  3007.5  337.9
638  213.5  51.3  2958.4  65.9  553.5  1715.7  624.2
639  93.6  25.8  3053.1  13.3  275.4  2632.2  332.5
640  117.7  58.8  1600.0  35.5  232.4  998.5  339.6

STATE_UT_NAME    DISTRICT
0  ANDAMAN And NICOBAR ISLANDS    NICOBAR
1  ANDAMAN And NICOBAR ISLANDS    SOUTH ANDAMAN
2  ANDAMAN And NICOBAR ISLANDS    N & M ANDAMAN
3  ARUNACHAL PRADESH    LOHIT
4  ARUNACHAL PRADESH    EAST SIANG

636  KERALA    IDUKKI
637  KERALA    KASARGOD
638  KERALA    PATHANAMTHITTA
639  KERALA    MAYANAD
640  LAKSHADWEEP    LAKSHADWEEP

[645 rows x 19 columns]
```

```
In [165.] # Printing the first 5 rows from data set:
rainfall_dataset.head()
```

#	Column	Non-Null Count	Dtype
0	JAN	641 non-null	float64
1	FEB	641 non-null	float64
2	MAR	641 non-null	float64
3	APR	641 non-null	float64
4	MAY	641 non-null	float64
5	JUN	641 non-null	float64
6	JUL	641 non-null	float64
7	AUG	641 non-null	float64
8	SEP	641 non-null	float64
9	OCT	641 non-null	float64
10	NOV	641 non-null	float64
11	DEC	641 non-null	float64
12	ANNUAL	641 non-null	float64
13	Jan-Feb	641 non-null	float64
14	Mar-May	641 non-null	float64
15	Jun-Sep	641 non-null	float64

```
In [166.] # Shape of the data:
rainfall_dataset.shape
```

```
Out[166.] (645, 19)
```

```
In [167.] # Getting information about the data:
rainfall_dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 645 entries, 0 to 640
Data columns (total 19 columns):
0      JAN      float64
1      FEB      float64
2      MAR      float64
3      APR      float64
4      MAY      float64
5      JUN      float64
6      JUL      float64
7      AUG      float64
8      SEP      float64
9      OCT      float64
10     NOV      float64
11     DEC      float64
12     ANNUAL    float64
13     Jan-Feb    float64
14     Mar-May    float64
15     Jun-Sep    float64
16     Oct-Dec    float64
17     STATE_UT_NAME  object
18     DISTRICT    object
dtypes: float64(17), object(2)
memory usage: 831.2 KB
```

```
In [168.] # Checking for missing values:
rainfall_dataset.isnull().sum()
```

Out[168.]

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec	STATE_UT_NAME	DISTRICT
count	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000
mean	18.35970	20.964399	30.034789	45.543214	81.535011	196.007332	332.033697	291.152262	194.609048	90.446334	31.717473	18.150568	1346.969679	39.339470	157.113105	101.333470	157.113105	101.333470	101.333470
std	11.202806	27.792996	45.451082	71.555279	111.960390	196.555284	221.364643	152.647325	99.303040	74.990685	59.371274	32.711009	838.878874	47.212773	213.445886	61.213444	213.445886	61.213444	61.213444
min	0.000000	0.000000	0.000000	0.000000	0.900000	3.800000	11.600000	14.100000	8.600000	3.100000	1.200000	0.000000	94.600000	0.000000	1.500000	1.500000	1.500000	1.500000	1.500000
25%	6.900000	7.000000	7.000000	5.000000	12.000000	68.800000	200.400000	194.600000	128.800000	34.300000	6.600000	5.300000	830.400000	14.700000	27.800000	6.200000	27.800000	6.200000	6.200000
50%	13.300000	12.300000	12.700000	15.100000	33.900000	131.900000	293.700000	284.800000	181.300000	62.600000	12.800000	7.900000	1116.200000	27.700000	67.200000	8.700000	67.200000	8.700000	8.700000
75%	19.200000	24.100000	33.200000	48.300000	91.900000	226.600000	374.800000	358.100000	234.100000	130.200000	32.300000	14.900000	1530.900000	41.100000	172.400000	11.700000	172.400000	11.700000	11.700000
max	114.500000	229.600000	387.900000	554.400000	733.700000	1476.200000	1820.900000	1522.100000	820.300000	517.700000	475.100000	297.700000	7229.300000	335.300000	1256.500000	52.000000	1256.500000	52.000000	52.000000

## Data Visualization

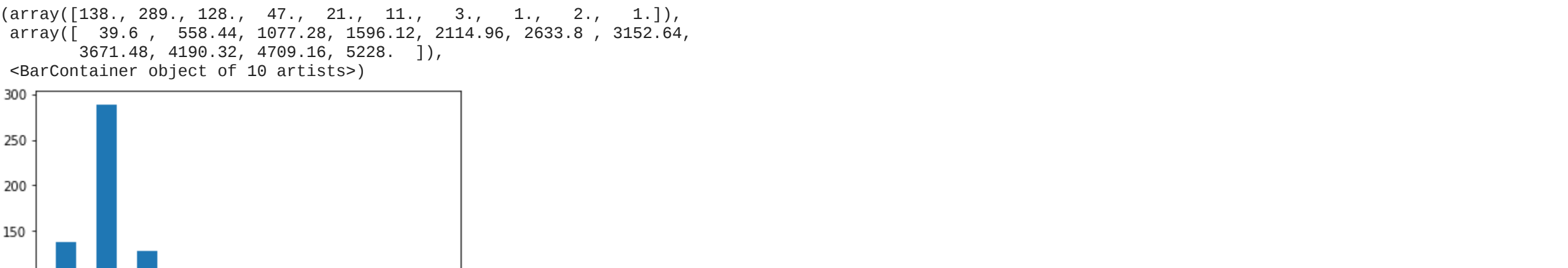
```
In [170.] # Rainfall in January-February:
plt.hist(rainfall_dataset['Jan-Feb'],width= 0.5)
```



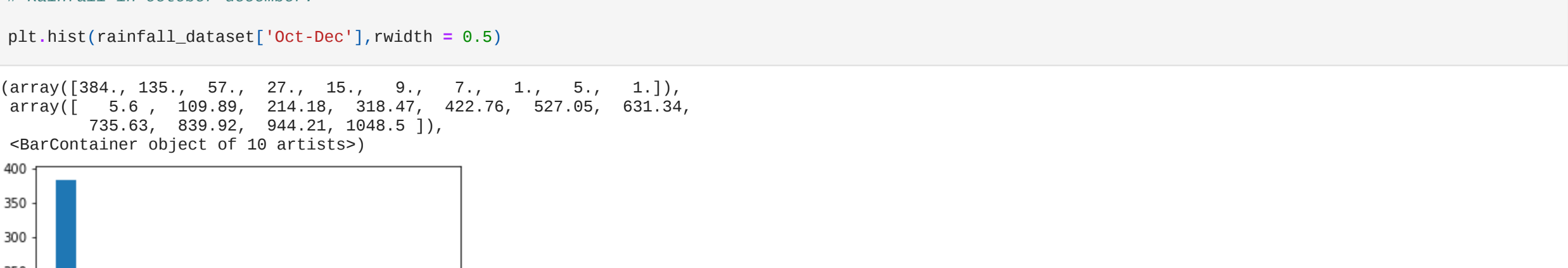
```
In [171.] # Rainfall in march-may:
plt.hist(rainfall_dataset['Mar-May'],width= 0.5)
```



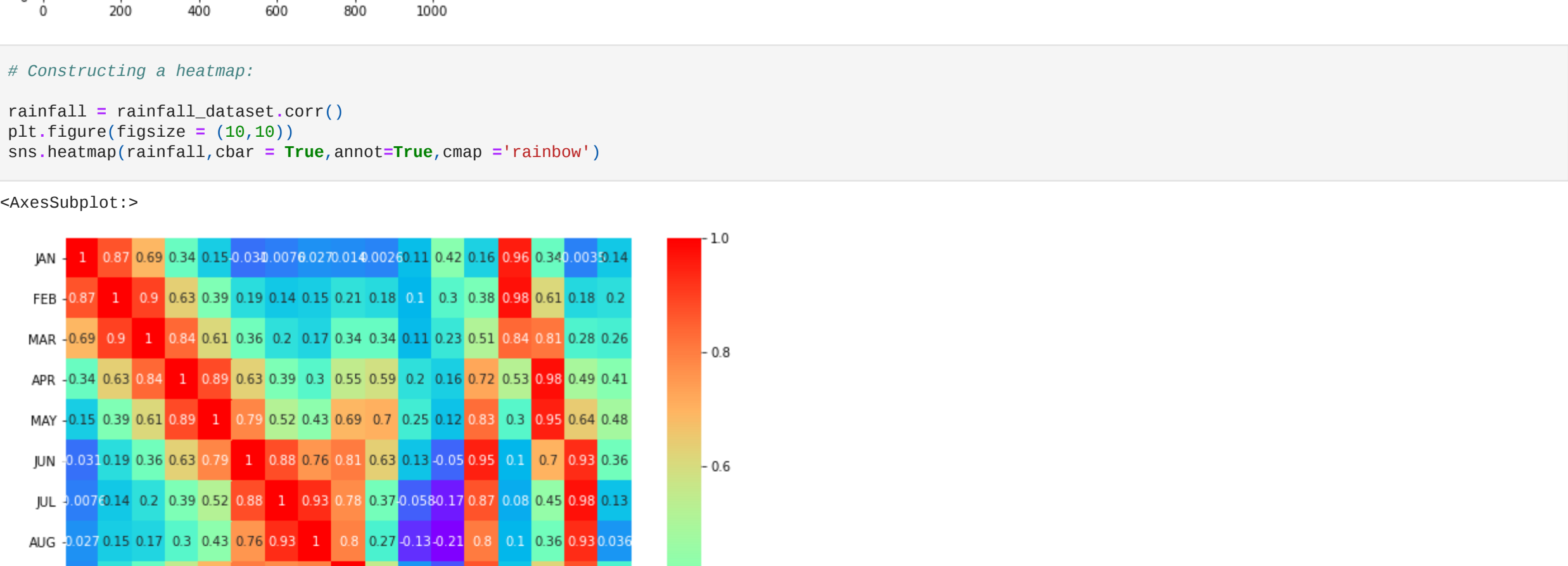
```
In [172.] # Rainfall in june-september:
plt.hist(rainfall_dataset['Jun-Sep'],width= 0.5)
```



```
In [173.] # Rainfall in october-december:
plt.hist(rainfall_dataset['Oct-Dec'],width= 0.5)
```



```
In [174.] # Constructing a heatmap:
rainfall = rainfall_dataset.corr()
plt.figure(figsize = (10,10))
sns.heatmap(rainfall,cbar = True,annot=True,cmap = 'rainbow')
```



## Preprocessing

```
In [176.] # Separating features and target:
X = rainfall_dataset.iloc[:,0:4].values
y = rainfall_dataset.iloc[:,4].values
```

```
In [176.] print(X)
```

[[[107.3 57.9 65.2 117. ]  
[ 43.7 26. 18.6 90.5]  
[ 32.7 15.9 8.6 53.4]  
[ 42.2 80.8 17.4 39.5]  
[ 33.3 79.5 105.9 216.5]  
[ 20.8 54.7 11.8 48.9]]]

```
In [177.] print(y)
```

[6.530e+01 1.880e+01 8.680e+00 1.764e+02 1.053e+02 8.530e+01 1.410e+02  
1.744e+02 3.835e+02 3.679e+02 1.661e+02 9.240e+01 3.626e+02 7.600e+01  
3.835e+02 5.890e+01 9.850e+01 5.890e+01 1.282e+02 1.683e+02 5.350e+01  
1.980e+01 2.620e+01 7.670e+01 5.580e+02 4.810e+01 7.530e+01 5.400e+01  
4.600e+01 1.139e+02 8.610e+01 1.697e+02 4.580e+01 5.080e+01 6.520e+01  
1.121e+02 1.238e+02 7.670e+01 5.280e+01 5.250e+01 5.580e+01 5.250e+01  
5.400e+01 5.750e+01 5.160e+01 5.670e+01 1.297e+02 1.151e+02 6.400e+01  
5.580e+01 4.870e+01 6.130e+01 4.870e+01 8.450e+01 4.750e+01 2.245e+02  
8.150e+01 5.690e+01 4.310e+01 1.224e+02 6.380e+01 7.270e+01 1.077e+02  
6.570e+01 5.580e+01 6.570e+01 6.570e+01 5.580e+01 1.184e+02 7.850e+01  
7.210e+01 1.655e+02 3.220e+01 5.770e+01 3.970e+01 1.450e+01 1.900e+01  
8.080e+00 1.950e+02 6.170e+01 1.950e+02 6.170e+01 2.280e+01 2.230e+01  
1.980e+01 1.315e+02 8.800e+01 2.460e+01 1.980e+01 2.110e+01 3.030e+01  
1.980e+01 3.180e+01 3.500e+01 3.500e+01 3.540e+01 2.610e+01 2.980e+01  
2.890e+01 3.130e+01 3.350e+01 1.530e+01 3.780e+01 1.630e+01 4.780e+01  
1.890e+01 2.610e+01 1.890e+01 3.780e+01 1.630e+01 3.780e+01 2.510e+01  
1.980e+01 1.640e+01 1.440e+01 2.690e+01 2.690e+01 8.510e+01 1.850e+01  
2.010e+01 3.850e+01 6.420e+01 3.530e+01 2.980e+01 1.850e+01 1.830e+01  
1.150e+01 1.720e+01 1.680e+01 1.680e+01 2.170e+01 1.260e+01 2.310e+01  
9.800e+00 1.370e+01 1.690e+01 1.910e+01 2.070e+01 1.120e+01 1.380e+01  
1.480e+01 2.070e+01 1.680e+01 2.280e+01 1.650e+01 2.040e+01 1.250e+01  
2.420e+01 1.130e+01 1.620e+01 7.880e+00 1.230e+01 8.780e+00 1.000e+01  
7.380e+00 1.280e+01 1.290e+01 7.880e+00 9.280e+00 7.880e+00 1.320e+01  
8.080e+00 9.180e+00 9.680e+00 9.280e+00 7.780e+00 8.580e+00 1.230e+01  
9.800e+00 9.180e+00 1.980e+01 9.980e+00 9.780e+00 7.280e+00 7.880e+00  
1.540e+01 1.540e+01 1.170e+01 1.080e+01 9.080e+00 1.750e+01 5.250e+01  
1.270e+01 1.080e+01 9.980e+00 5.280e+00 1.280e+01 7.680e+00 8.080e+00  
1.230e+01 6.780e+00 6.280e+00 8.280e+00 1.080e+01 0.280e+00 7.880e+00  
9.980e+00 7.380e+00 8.080e+00 5.280e+00 1.080e+01 5.580e+00 6.980e+00  
1.980e+00 1.130e+01 7.380e+00 1.630e+01 1.080e+01 5.880e+00 1.180e+01  
1.280e+01 7.480e+00 9.980e+00 1.530e+01 1.280e+01 9.880e+00 9.380e+00  
1.980e+01 8.580e+00 9.980e+00 8.280e+00 8.280e+00 8.680e+00 5.580e+00  
1.150e+01 1.650e+01 1.180e+01 1.250e+01 1.250e+01 1.650e+01 7.250e+00  
2.410e+01 1.170e+01 1.310e+01 9.980e+00 8.980e+00 5.380e+00 5.780e+00  
5.380e+00 6.580e+00 8.280e+00 9.980e+00 1.680e+01 1.380e+01 1.440e+01  
1.380e+01 1.920e+01 1.920e+01 2.130e+01 1.380e+01 1.880e+01 5.880e+00  
1.640e+00 1.130e+01 6.280e+00 1.630e+01 1.630e+01 5.880e+00 1.180e+01  
2.130e+01 7.480e+00 7.480e+00 9.980e+00 1.530e+01 1.280e+01 9.880e+00  
1.980e+01 8.580e+00 9.980e+00 8.280e+00 8.280e+00 8.680e+00 5.580e+00  
1.150e+01 1.650e+01 1.180e+01 1.250e+01 1.250e+01 1.650e+01 7.250e+00  
2.410e+01 1.170e+01 1.310e+01 9.980e+00 8.980e+00 5.380e+00 5.780e+00  
5.380e+00 6.580e+00 8.280e+00 9.980e+00 1.680e+01 1.380e+01 1.440e+01  
1.380e+01 1.920e+01 1.920e+01 2.130e+01 1.380e+01 1.880e+01 5.880e+00  
1.640e+00 1.130e+01 6.280e+00 1.630e+01 1.630e+01 5.880e+00 1.180e+01  
2.130e+01 7.480e+00 7.480e+00 9.980e+00 1.530e+01 1.280e+01 9.880e+00  
1.980e+01 8.580e+00 9.980e+00 8.280e+00 8.280e+00 8.680e+00 5.580e+00  
1.150e+01 1.650e+01 1.180e+01 1.250e+01 1.250e+01 1.650e+01 7.250e+00  
2.410e+01 1.170e+01 1.310e+01 9.980e+00 8.980e+00 5.380e+00 5.780e+00  
5.380e+00 6.580e+00 8.280e+00 9.980e+00 1.680e+01 1.380e+01 1.440e+01  
1.380e+01 1.920e+01 1.920e+01 2.130e+01 1.380e+01 1.880e+01 5.880e+00  
1.640e+00 1.130e+01 6.280e+00 1.630e+01 1.630e+01 5.880e+00 1.180e+01  
2.130e+01 7.480e+00 7.480e+00 9.980e+00 1.530e+01 1.280e+01 9.880e+00  
1.980e+01 8.580e+00 9.980e+00 8.280e+00 8.280e+00 8.680e+00 5.580e+00  
1.150e+01 1.650e+01 1.180e+01 1.250e+01 1.250e+01 1.650e+01 7.250e+00  
2.410e+01 1.170e+01 1.310e+01 9.980e+00 8.980e+00 5.380e+00 5.780e+00  
5.380e+00 6.580e+00 8.280e+00 9.980e+00 1.680e+01 1.380e+01 1.440e+01  
1.380e+01 1.920e+01 1.920e+01 2.130e+01 1.380e+01 1.880e+01 5.880e+00  
1.640e+00 1.130e+01 6.280e+00 1.630e+01 1.630e+01 5.880e+00 1.180e+01  
2.130e+01 7.480e+00 7.480e+00 9.980e+00 1.530e+01 1.280e+01 9.880e+00  
1.980e+01 8.580e+00 9.980e+00 8.280e+00 8.280e+00 8.680e+00 5.580e+00  
1.150e+01 1.650e+01 1.180e+01 1.250e+01 1.250e+01 1.650e+01 7.250e+00  
2.410e+01 1.170e+01 1.310e+01 9.980e+00 8.980e+00 5.380e+00 5.780e+00  
5.380e+00 6.580e+00 8.280e+00 9.980e+00 1.680e+01 1.380e+01 1.440e+01  
1.380e+01 1.920e+01 1.920e+01 2.130e+01 1.380e+01 1.880e+01 5.880e+00  
1.640e+00 1.130e+01 6.280e+00 1.630e+01 1.630e+01 5.880e+00 1.180e+01  
2.130e+01 7.480e+00 7.480e+00 9.980e+00 1.530e+01 1.280e+01 9.880e+00  
1.980e+01 8.580e+00 9.980e+00 8.280e+00 8.280e+00 8.680e+00 5.580e+00  
1.150e+01 1.650e+01 1.180e+01 1.250e+01 1.250e+01 1.650e+01 7.250e+00  
2.410e+01 1.170e+01 1.310e+01 9.980e+00 8.980e+00 5.380e