Rainfall Analysis and Prediction

October 9, 2018

India Rainfall Analysis

```
In [1]: import numpy as np # linear algebra
        import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
        import matplotlib.pyplot as plt
        from subprocess import check_output
        import seaborn as sns
```

```
In [2]: data = pd.read_csv("../data/rainfall_in_india_1901-2015.csv",sep=",")
        data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4116 entries, 0 to 4115
Data columns (total 19 columns):
SUBDIVISION    4116 non-null object
YEAR           4116 non-null int64
JAN            4112 non-null float64
FEB            4113 non-null float64
MAR            4110 non-null float64
APR            4112 non-null float64
MAY            4113 non-null float64
JUN            4111 non-null float64
JUL            4109 non-null float64
AUG            4112 non-null float64
SEP            4110 non-null float64
OCT            4109 non-null float64
NOV            4105 non-null float64
DEC            4106 non-null float64
ANNUAL         4090 non-null float64
Jan-Feb        4110 non-null float64
Mar-May        4107 non-null float64
Jun-Sep        4106 non-null float64
Oct-Dec        4103 non-null float64
dtypes: float64(17), int64(1), object(1)
memory usage: 611.0+ KB
```

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

```
Out[3]:                  SUBDIVISION  YEAR   JAN    FEB   MAR    APR    MAY    JUN  \
        0  ANDAMAN & NICOBAR ISLANDS  1901  49.2   87.1  29.2    2.3  528.8  517.5
        1  ANDAMAN & NICOBAR ISLANDS  1902   0.0  159.8  12.2    0.0  446.1  537.1
        2  ANDAMAN & NICOBAR ISLANDS  1903  12.7  144.0   0.0    1.0  235.1  479.9
        3  ANDAMAN & NICOBAR ISLANDS  1904   9.4   14.7   0.0  202.4  304.5  495.1
```

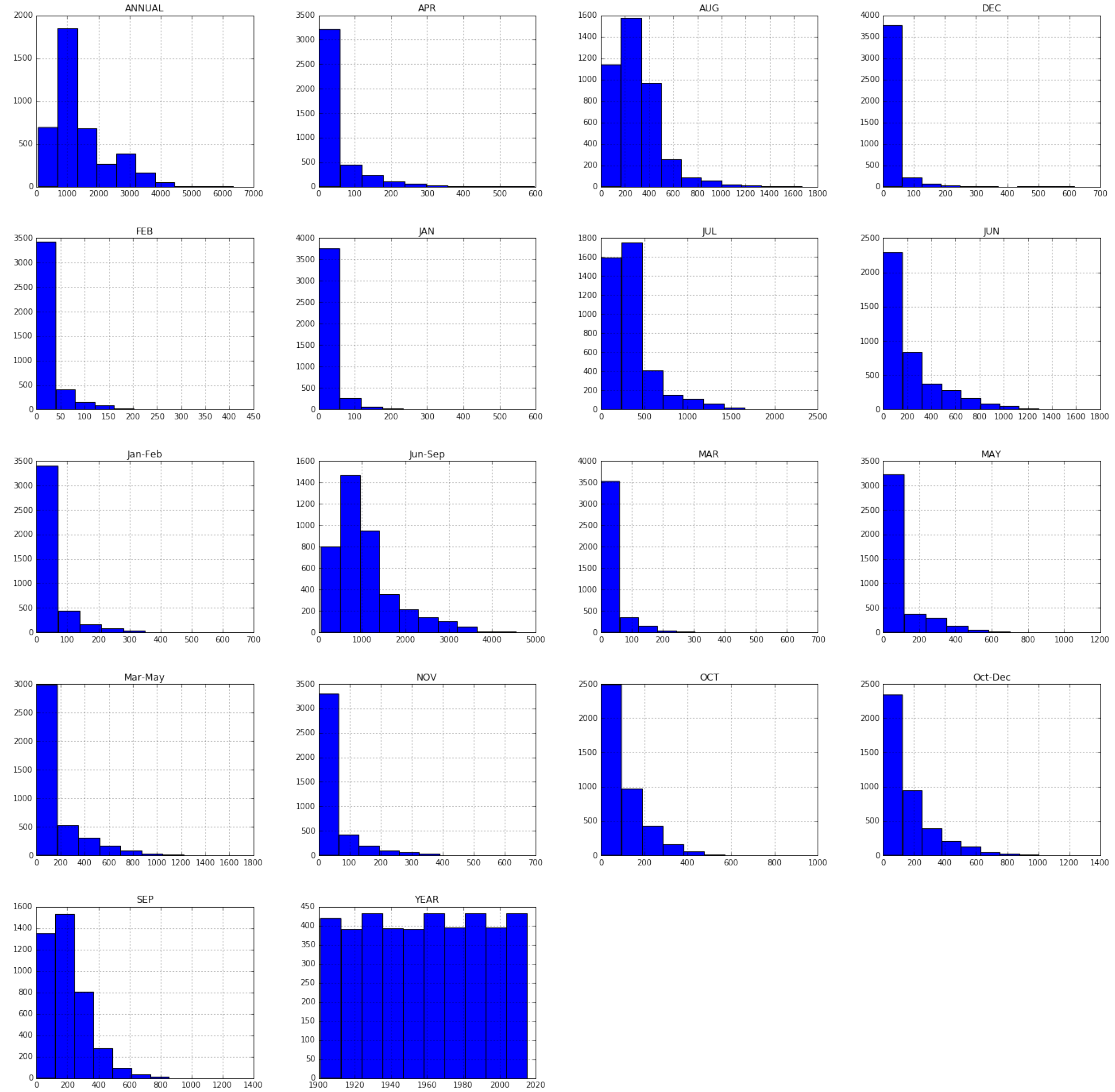
```
4  ANDAMAN & NICOBAR ISLANDS  1905   1.3    0.0   3.3   26.9  279.5  628.7

     JUL    AUG    SEP    OCT    NOV    DEC  ANNUAL  Jan-Feb  Mar-May  \
0  365.1  481.1  332.6  388.5  558.2   33.6  3373.2    136.3    560.3
1  228.9  753.7  666.2  197.2  359.0  160.5  3520.7    159.8    458.3
2  728.4  326.7  339.0  181.2  284.4  225.0  2957.4    156.7    236.1
3  502.0  160.1  820.4  222.2  308.7   40.1  3079.6     24.1    506.9
4  368.7  330.5  297.0  260.7   25.4  344.7  2566.7      1.3    309.7

   Jun-Sep  Oct-Dec
0   1696.3    980.3
1   2185.9    716.7
2   1874.0    690.6
3   1977.6    571.0
4   1624.9    630.8
```

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

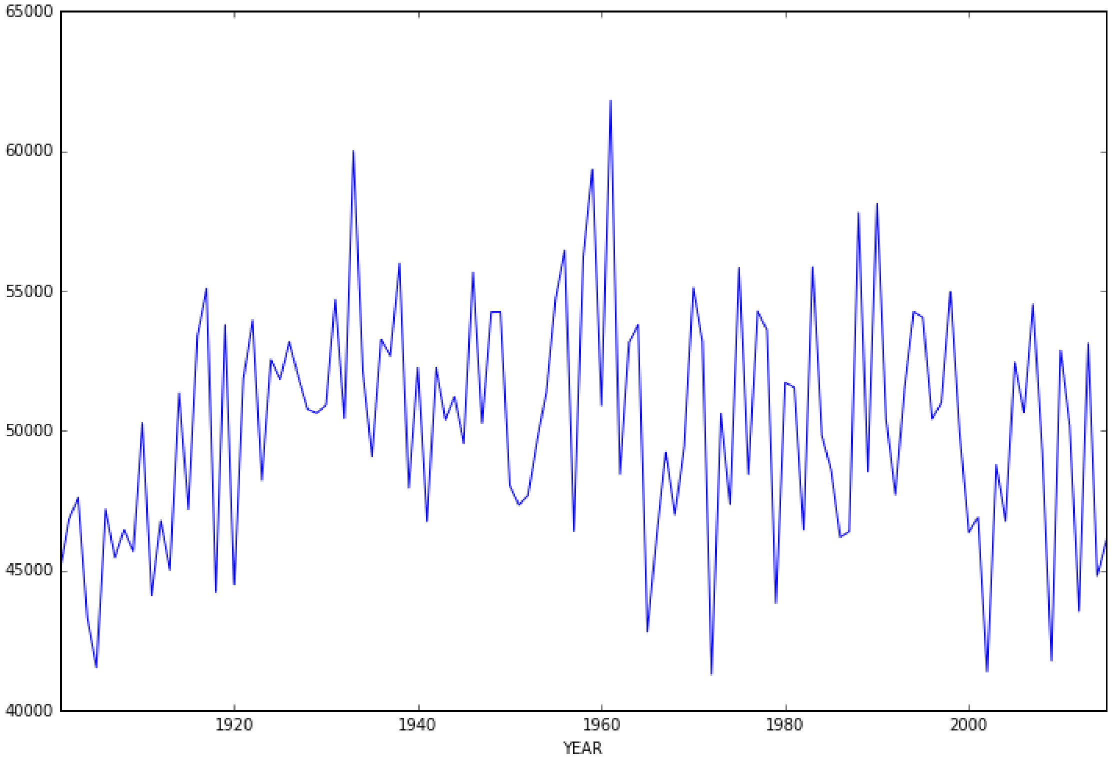
```
Out[4]:               YEAR          JAN          FEB          MAR          APR  \
        count  4116.000000  4112.000000  4113.000000  4110.000000  4112.000000
        mean   1958.218659    18.957320    21.805325    27.359197    43.127432
        std      33.140898    33.585371    35.909488    46.959424    67.831168
        min    1901.000000     0.000000     0.000000     0.000000     0.000000
        25%    1930.000000     0.600000     0.600000     1.000000     3.000000
        50%    1958.000000     6.000000     6.700000     7.800000    15.700000
        75%    1987.000000    22.200000    26.800000    31.300000    49.950000
        max    2015.000000   583.700000   403.500000   605.600000   595.100000

                       MAY          JUN          JUL          AUG          SEP  \
        count  4113.000000  4111.000000  4109.000000  4112.000000  4110.000000
        mean     85.745417   230.234444   347.214334   290.263497   197.361922
        std     123.234904   234.710758   269.539667   188.770477   135.408345
        min       0.000000     0.400000     0.000000     0.000000     0.100000
        25%       8.600000    70.350000   175.600000   155.975000   100.525000
        50%      36.600000   138.700000   284.800000   259.400000   173.900000
        75%      97.200000   305.150000   418.400000   377.800000   265.800000
        max    1168.600000  1609.900000  2362.800000  1664.600000  1222.000000

                       OCT          NOV          DEC       ANNUAL      Jan-Feb  \
        count  4109.000000  4105.000000  4106.000000  4090.000000  4110.000000
        mean     95.507009    39.866163    18.870580  1411.008900    40.747786
        std      99.519134    68.685410    42.369611   903.846565    59.308277
        min       0.000000     0.000000     0.000000    62.300000     0.000000
        25%      14.600000     0.700000     0.100000   804.500000     4.100000
        50%      65.200000     9.500000     3.000000  1121.300000    19.200000
        75%     148.400000    46.100000    17.500000  1644.775000    50.375000
        max     948.300000   648.900000   617.500000  6331.100000   699.500000

                   Mar-May      Jun-Sep      Oct-Dec
        count  4107.000000  4106.000000  4103.000000
        mean    155.901753  1064.724769   154.100487
        std     201.316965   707.741531   166.942660
        min       0.000000    57.400000     0.000000
        25%      24.050000   573.850000    34.200000
        50%      74.800000   881.100000    98.200000
```

```
75%      196.950000  1288.175000   213.500000
max     1745.800000  4536.900000  1252.500000
```

```
In [5]: data.hist(figsize=(24,24));
```



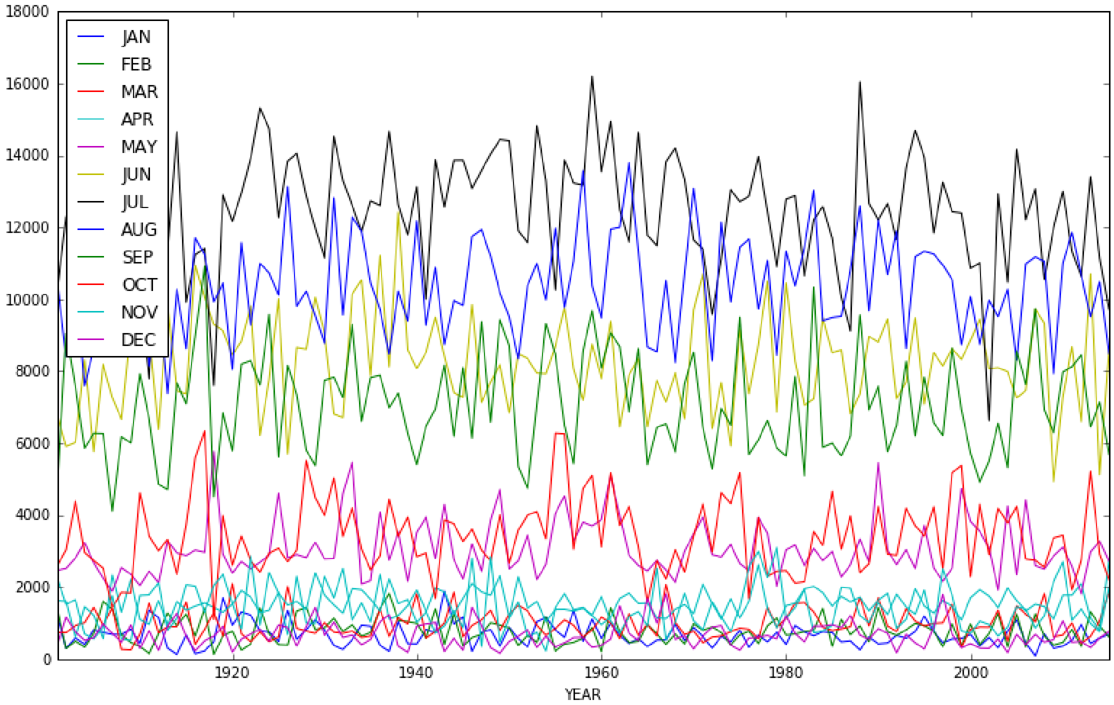
```
In [6]: data.groupby("YEAR").sum()['ANNUAL'].plot(figsize=(12,8));
```



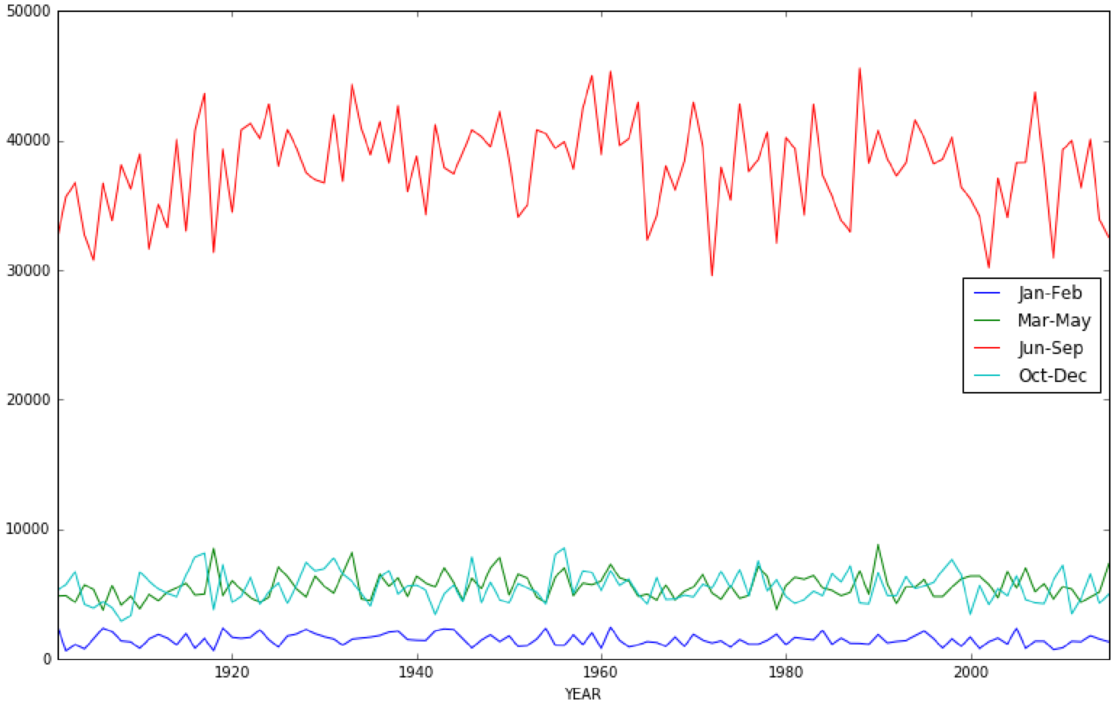
```
In [7]: data.columns
```

```
Out[7]: Index([u'SUBDIVISION', u'YEAR', u'JAN', u'FEB', u'MAR', u'APR', u'MAY', u'JUN',
               u'JUL', u'AUG', u'SEP', u'OCT', u'NOV', u'DEC', u'ANNUAL', u'Jan-Feb',
               u'Mar-May', u'Jun-Sep', u'Oct-Dec'],
              dtype='object')
```

```
In [8]: data[['YEAR', 'JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL',
               'AUG', 'SEP', 'OCT', 'NOV', 'DEC']].groupby("YEAR").sum().plot(figsize=(13,8));
```

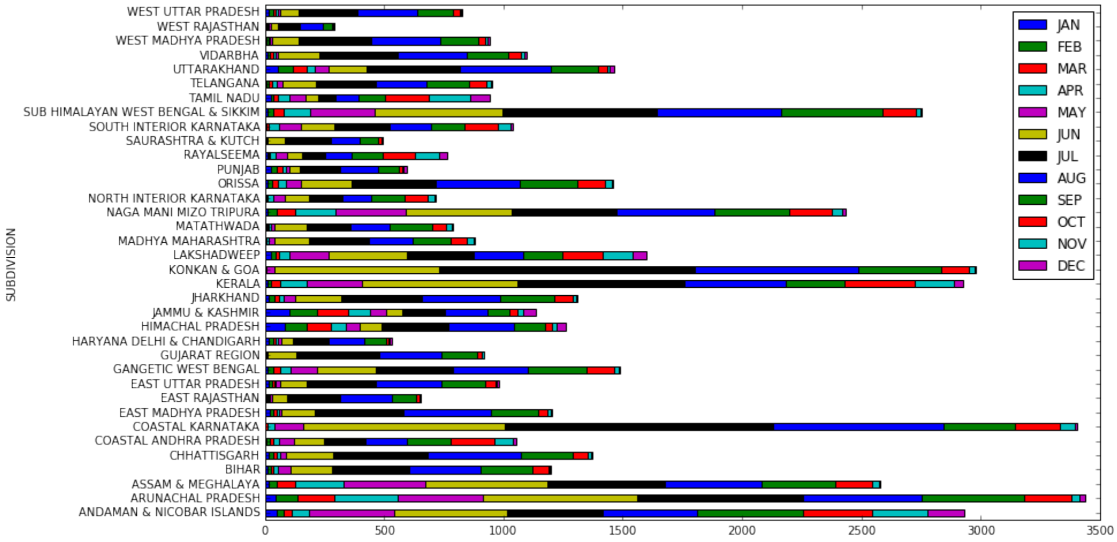


```
In [9]: data[['YEAR','Jan-Feb', 'Mar-May',
          'Jun-Sep', 'Oct-Dec']].groupby("YEAR").sum().plot(figsize=(13,8));
```



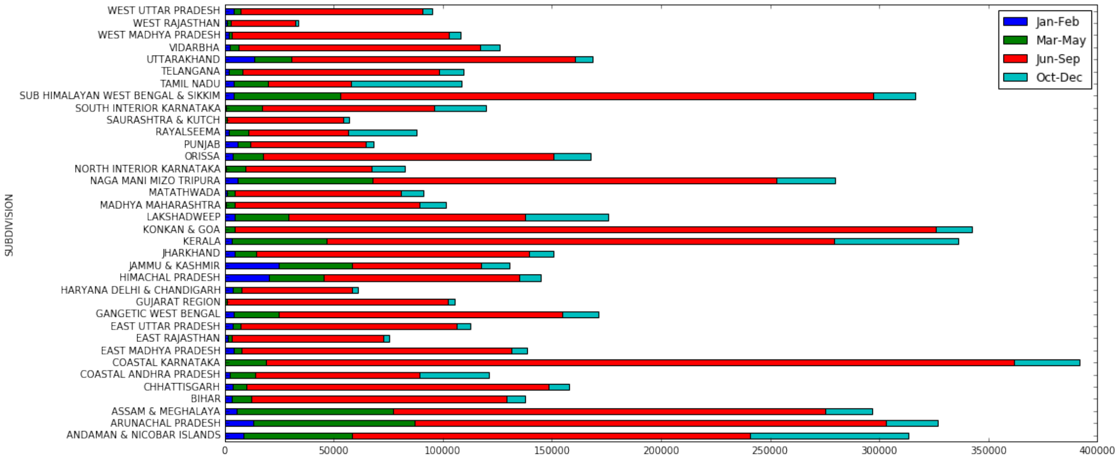
Span of Jun-Sep has maximum rainfall records.

```
In [10]: data[['SUBDIVISION', 'JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL',
               'AUG', 'SEP', 'OCT', 'NOV', 'DEC']].groupby("SUBDIVISION").mean().plot.barh(stacked=True
```

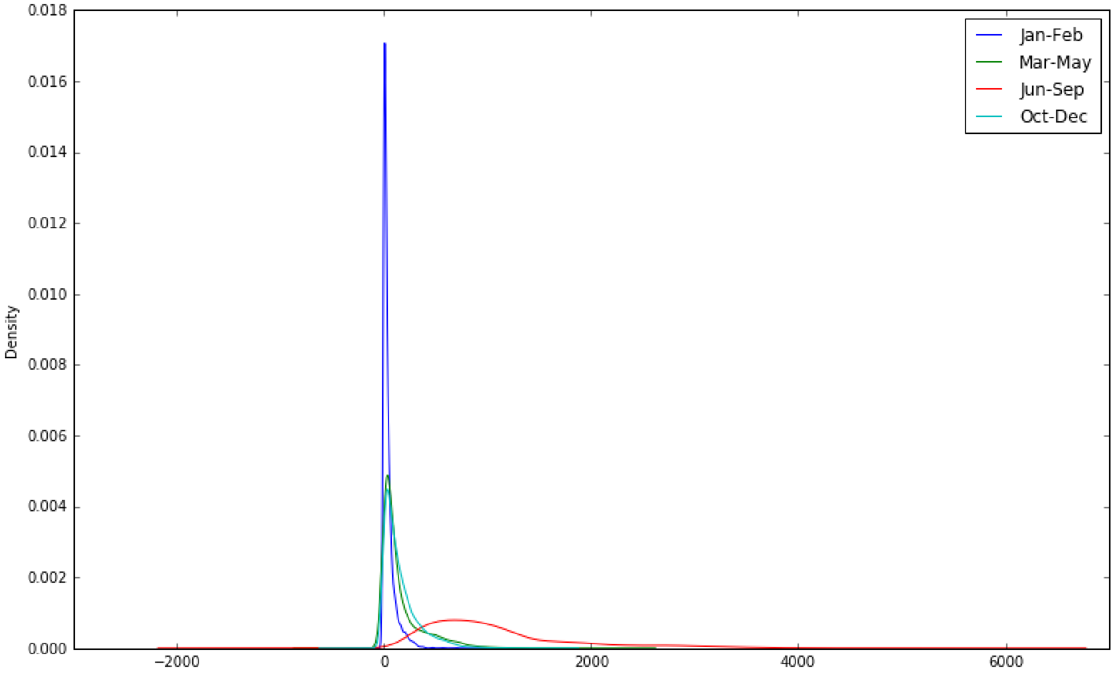


Graph shows top 3 subdivisions having high average rainfall: 1. ARUNACHAL PRADESH 2. COASTAL KARNATAKA 3. KOKAN & GOA

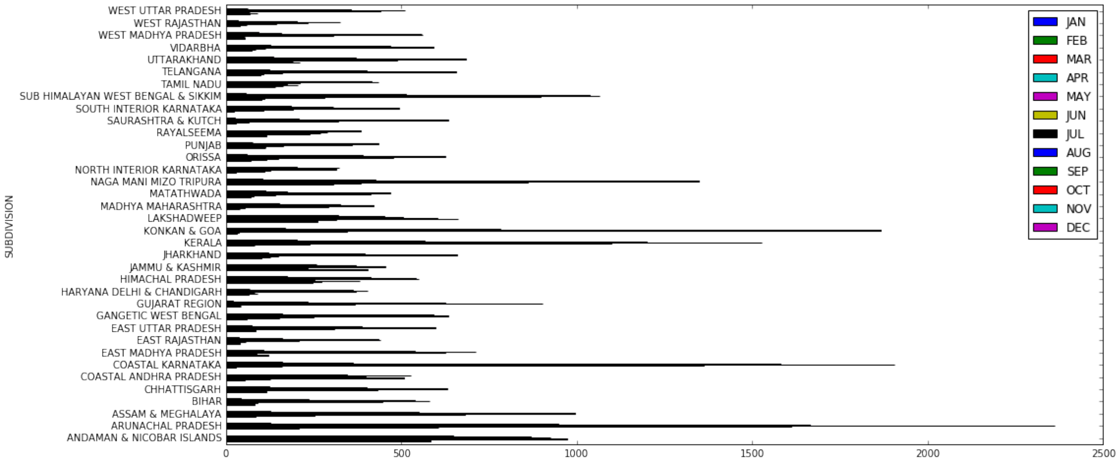
```
In [11]: data[['SUBDIVISION', 'Jan-Feb', 'Mar-May',
               'Jun-Sep', 'Oct-Dec']].groupby("SUBDIVISION").sum().plot.barh(stacked=True,figsize=(16,8
```



```
In [12]: data[['Jan-Feb', 'Mar-May',
               'Jun-Sep', 'Oct-Dec']].plot(kind="kde",figsize=(13,8));
```



```
In [13]: data[['SUBDIVISION', 'JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL',
                  'AUG', 'SEP', 'OCT', 'NOV', 'DEC']].groupby("SUBDIVISION").max().plot.barh(figsize=(16,8
```



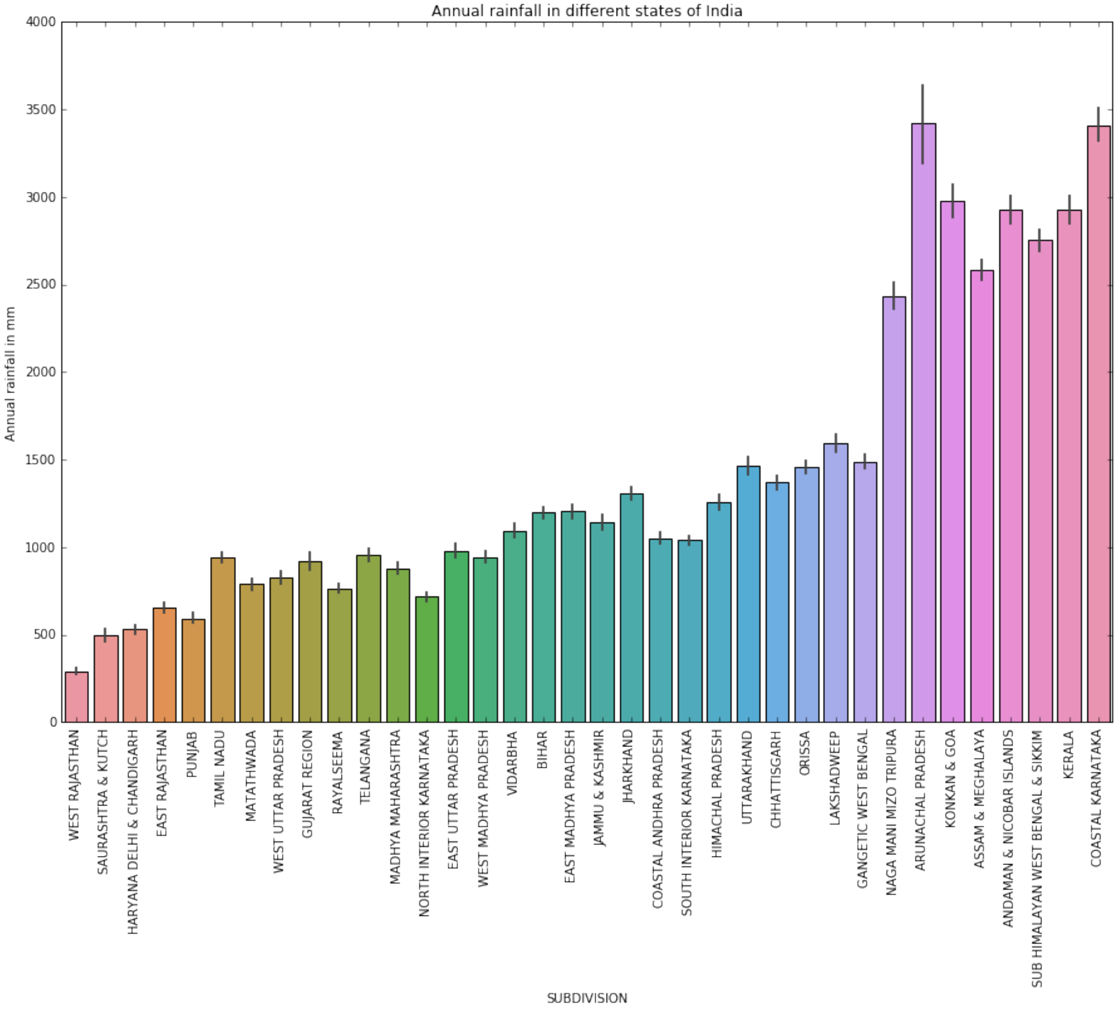
```
In [14]: S = data.sort_values('ANNUAL',ascending=True)
         S.head()
```

```
Out[14]:              SUBDIVISION  YEAR  JAN  FEB  MAR  APR   MAY   JUN   JUL   AUG  \
         1834      WEST RAJASTHAN  1918  0.4  0.0  2.7  0.5   0.7   5.6   5.1  40.8
         1918      WEST RAJASTHAN  2002  0.3  2.6  1.0  1.4  11.2  26.9   2.4  17.2
         2478  SAURASHTRA & KUTCH  1987  0.0  0.0  0.4  0.0   0.1  31.4  18.4  33.5
         1821      WEST RAJASTHAN  1905  6.3  4.8  0.7  1.3   0.3   4.9  30.1   0.6
```

```
2409   SAURASHTRA & KUTCH  1918  0.0  0.0  0.6  0.6   3.8   9.9  45.0  49.0

       SEP   OCT  NOV  DEC  ANNUAL  Jan-Feb  Mar-May  Jun-Sep  Oct-Dec
1834   5.9   0.2  0.2  0.1    62.3      0.4      4.0     57.4      0.5
1918  20.6   0.1  2.5  6.0    92.4      3.0     13.7     67.2      8.6
2478   2.1   0.1  3.2  3.3    92.7      0.0      0.5     85.5      6.6
1821  64.5   0.0  0.0  0.9   114.4     11.0      2.4    100.1      0.9
2409   8.6   0.2  0.0  0.0   117.6      0.0      5.0    112.4      0.2
```

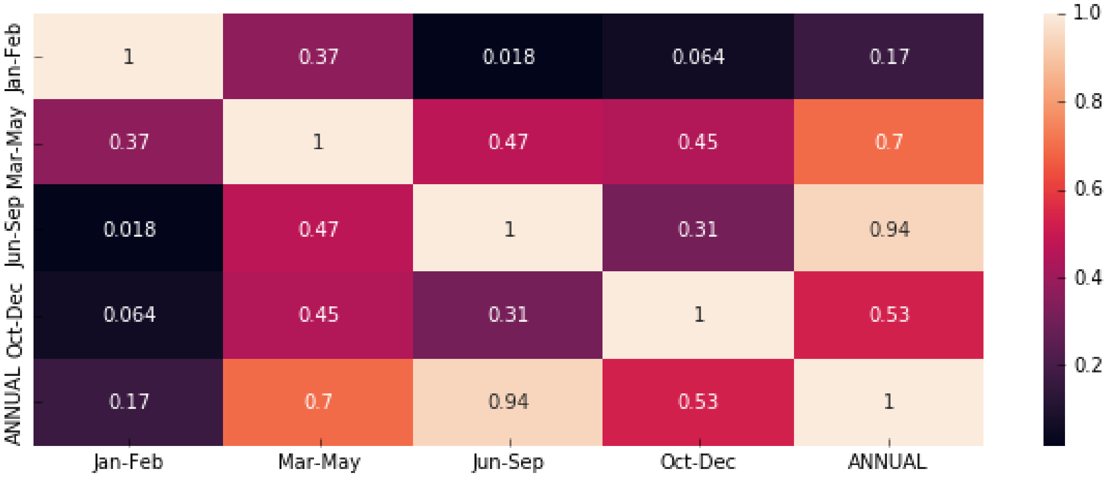
```
In [15]: fig = plt.figure(figsize=(15, 10))
         ax = plt.xticks(rotation=90)
         ax = sns.barplot(x='SUBDIVISION', y='ANNUAL', data=S)
         plt.xlabel('SUBDIVISION')
         plt.ylabel('Annual rainfall in mm')
         plt.title('Annual rainfall in different states of India')
```

```
Out[15]: <matplotlib.text.Text at 0x7efc70e193d0>
```

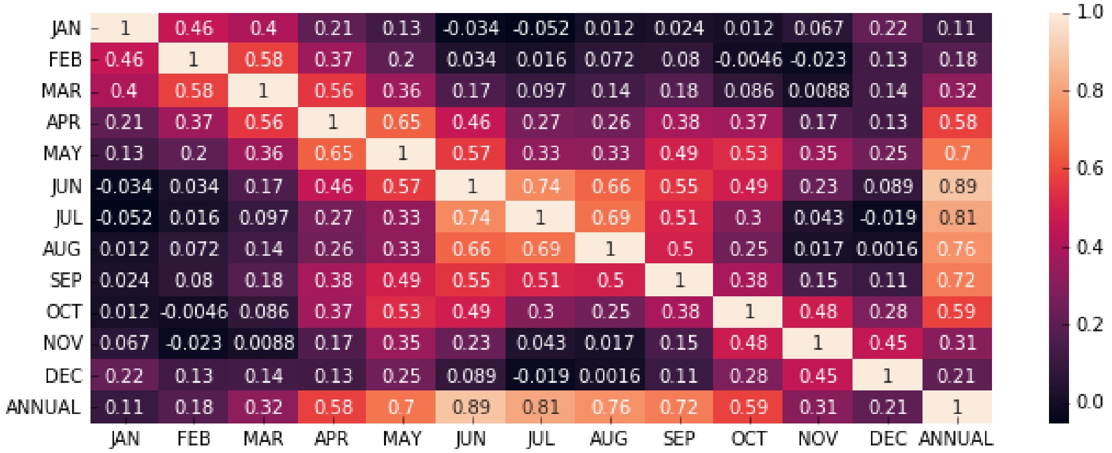


```
In [16]: plt.figure(figsize=(11,4))
         sns.heatmap(data[['Jan-Feb','Mar-May','Jun-Sep','Oct-Dec','ANNUAL']].corr(),annot=True)
         plt.show()
```

```
/home/sudheer.achary/.local/lib/python2.7/site-packages/pandas/core/computation/check.py:17: UserWarning
The minimum supported version is 2.4.6

  ver=ver, min_ver=_MIN_NUMEXPR_VERSION), UserWarning)
```



```
In [17]: plt.figure(figsize=(11,4))
         sns.heatmap(data[['JAN','FEB','MAR','APR','MAY','JUN','JUL','AUG','SEP','OCT','NOV','DEC','ANNU
         plt.show()
```



Graph shows Max rainfall for specific month registered in ARUNACHAL PRADESH.
District wise details

```
In [18]: district = pd.read_csv("../data/district_wise_rainfall_normal.csv",sep=",")
         district.info()
```

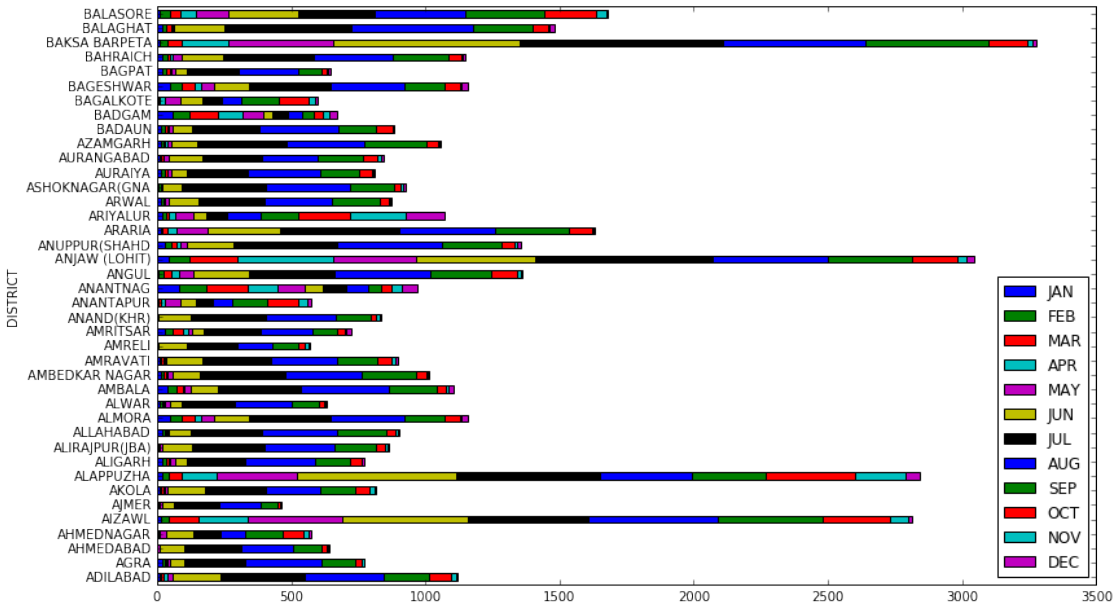
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 641 entries, 0 to 640
```

```
Data columns (total 19 columns):
STATE_UT_NAME    641 non-null object
DISTRICT         641 non-null object
JAN              641 non-null float64
FEB              641 non-null float64
MAR              641 non-null float64
APR              641 non-null float64
MAY              641 non-null float64
JUN              641 non-null float64
JUL              641 non-null float64
AUG              641 non-null float64
SEP              641 non-null float64
OCT              641 non-null float64
NOV              641 non-null float64
DEC              641 non-null float64
ANNUAL           641 non-null float64
Jan-Feb          641 non-null float64
Mar-May          641 non-null float64
Jun-Sep          641 non-null float64
Oct-Dec          641 non-null float64
dtypes: float64(17), object(2)
memory usage: 95.2+ KB
```

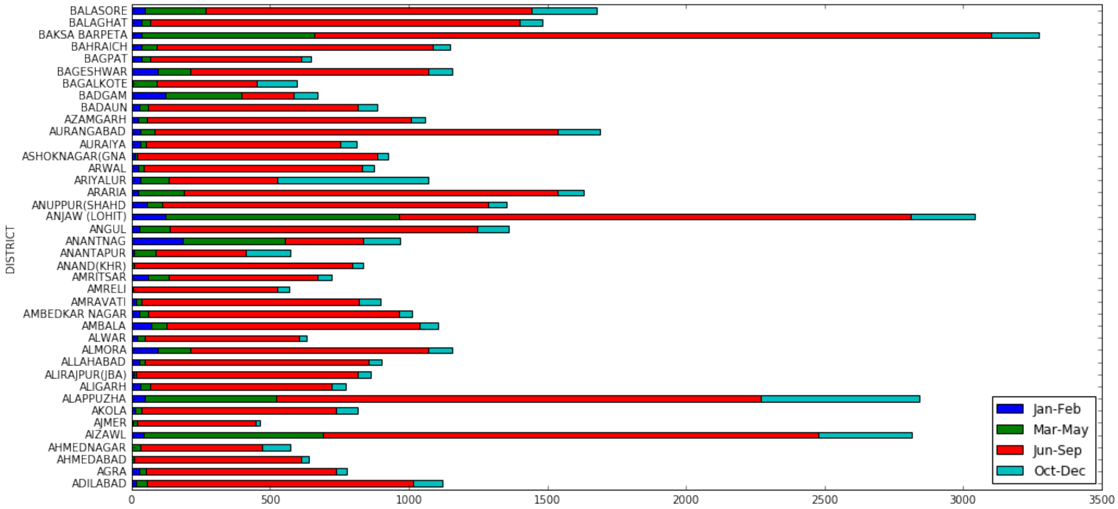
```
In [19]: district.head()
```

```
Out[19]:                 STATE_UT_NAME       DISTRICT    JAN   FEB    MAR    APR  \
         0  ANDAMAN And NICOBAR ISLANDS        NICOBAR  107.3  57.9   65.2  117.0
         1  ANDAMAN And NICOBAR ISLANDS  SOUTH ANDAMAN   43.7  26.0   18.6   90.5
         2  ANDAMAN And NICOBAR ISLANDS  N & M ANDAMAN   32.7  15.9    8.6   53.4
         3            ARUNACHAL PRADESH          LOHIT   42.2  80.8  176.4  358.5
         4            ARUNACHAL PRADESH     EAST SIANG   33.3  79.5  105.9  216.5

              MAY    JUN    JUL    AUG    SEP    OCT    NOV    DEC  ANNUAL  Jan-Feb  \
         0  358.5  295.5  285.0  271.9  354.8  326.0  315.2  250.9  2805.2    165.2
         1  374.4  457.2  421.3  423.1  455.6  301.2  275.8  128.3  3015.7     69.7
         2  343.6  503.3  465.4  460.9  454.8  276.1  198.6  100.0  2913.3     48.6
         3  306.4  447.0  660.1  427.8  313.6  167.1   34.1   29.8  3043.8    123.0
         4  323.0  738.3  990.9  711.2  568.0  206.9   29.5   31.7  4034.7    112.8

            Mar-May  Jun-Sep  Oct-Dec
         0    540.7   1207.2    892.1
         1    483.5   1757.2    705.3
         2    405.6   1884.4    574.7
         3    841.3   1848.5    231.0
         4    645.4   3008.4    268.1
```

```
In [20]: district[['DISTRICT', 'JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL',
               'AUG', 'SEP', 'OCT', 'NOV', 'DEC']].groupby("DISTRICT").mean()[:40].plot.barh(stacked=Tr
```



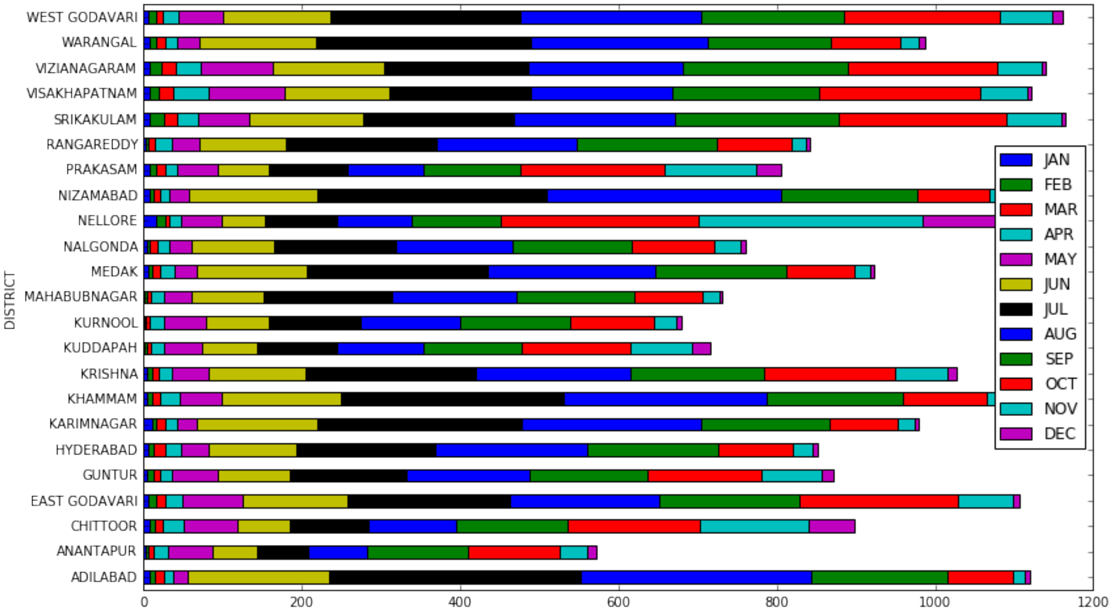
```
In [21]: district[['DISTRICT', 'Jan-Feb', 'Mar-May',
             'Jun-Sep', 'Oct-Dec']].groupby("DISTRICT").sum()[:40].plot.barh(stacked=True,figsize=(16
```



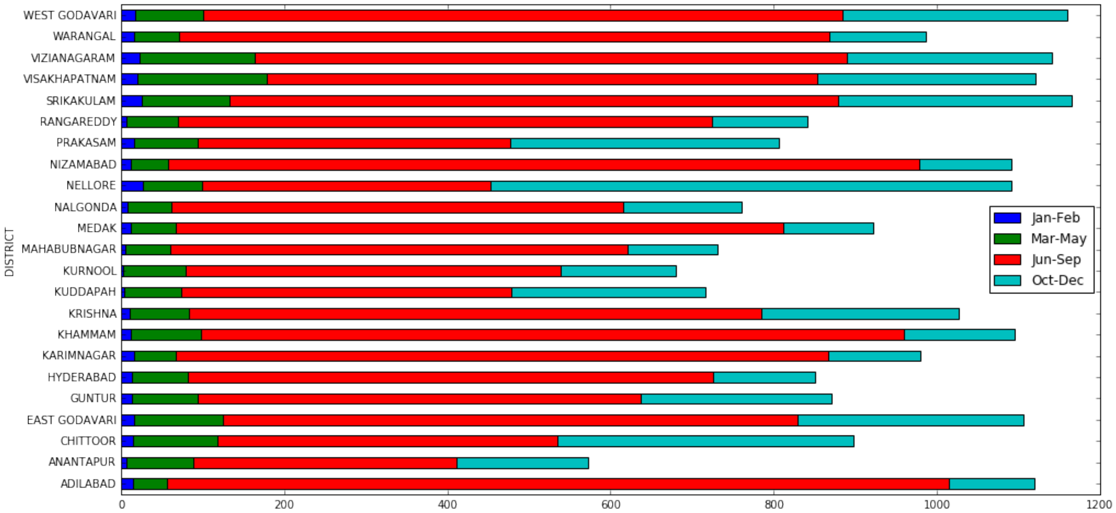
Andhra Pradesh Data

```
In [22]: ap_data = district[district['STATE_UT_NAME'] == 'ANDHRA PRADESH']
```

```
In [23]: ap_data[['DISTRICT', 'JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL',
             'AUG', 'SEP', 'OCT', 'NOV', 'DEC']].groupby("DISTRICT").mean()[:40].plot.barh(stacked=Tr
```



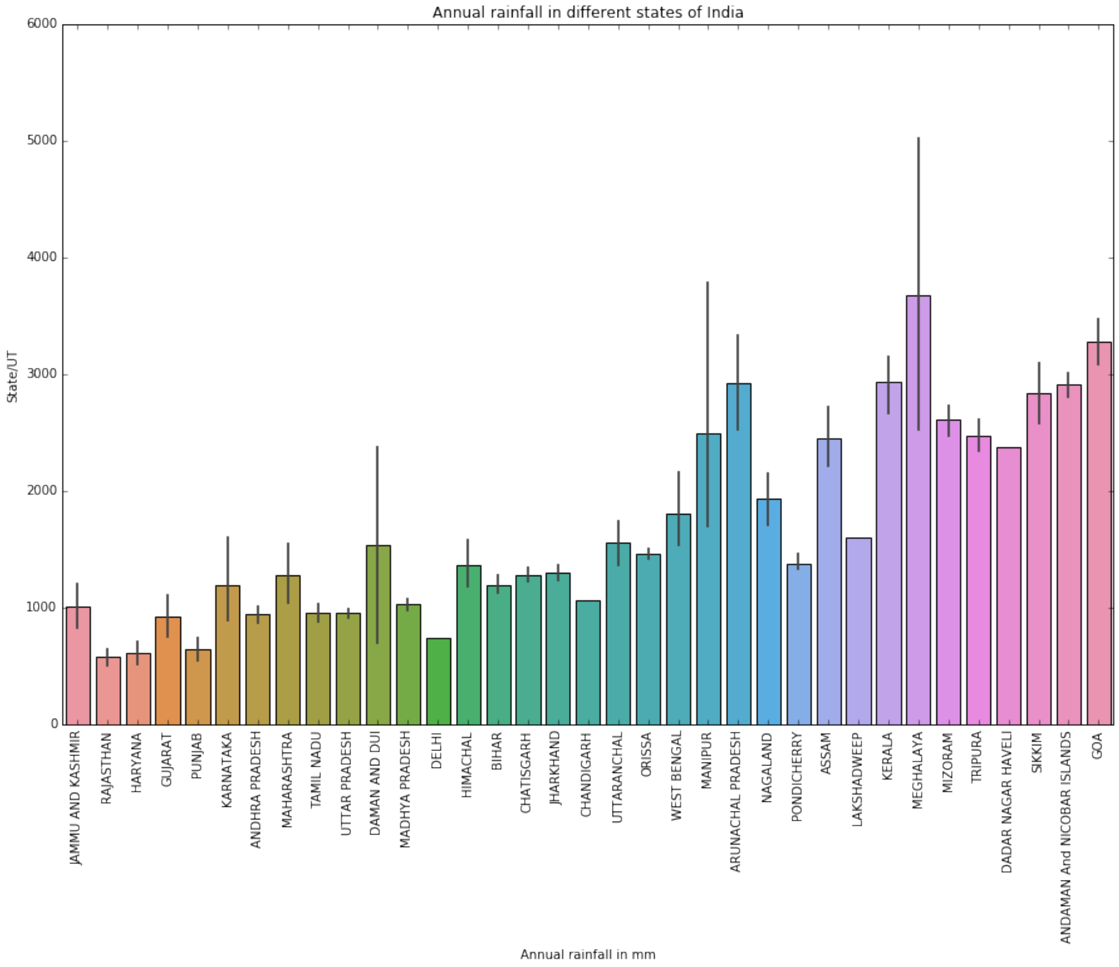
```
In [24]: ap_data[['DISTRICT', 'Jan-Feb', 'Mar-May',
                  'Jun-Sep', 'Oct-Dec']].groupby("DISTRICT").sum()[:40].plot.barh(stacked=True,figsize=(16
```



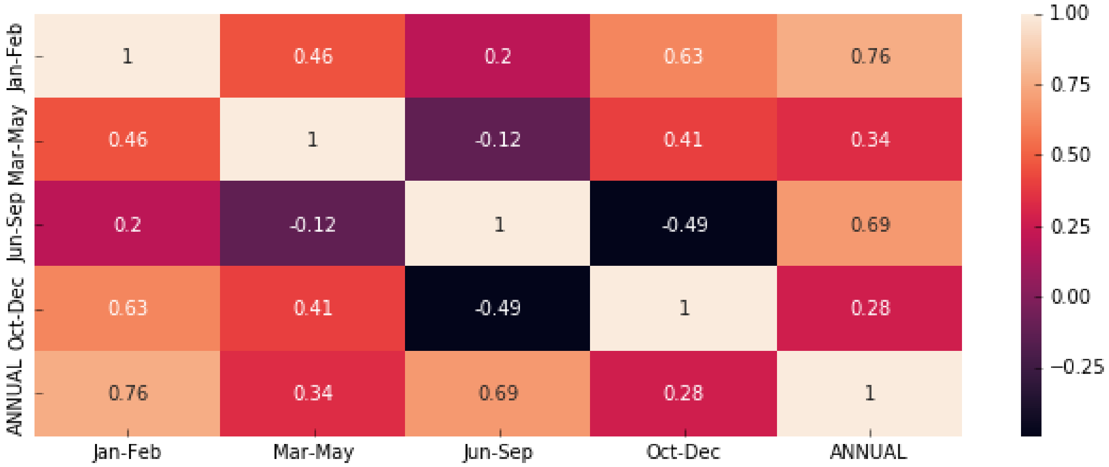
```
In [25]: S = district.sort_values('ANNUAL',ascending=True)
         S.head()

Out[25]:          STATE_UT_NAME        DISTRICT   JAN   FEB   MAR   APR   MAY   JUN  \
         353  JAMMU AND KASHMIR   LADAKH (LEH)   8.1   7.3   8.8   8.0   8.0   3.8
         374          RAJASTHAN      JAISALMER   1.7   2.9   2.2   3.8   6.6  19.4
         359  JAMMU AND KASHMIR         KARGIL  25.6  25.6  25.1  19.7  22.0   5.5
         373          RAJASTHAN  SRI GANGANAGA   6.0   8.0   9.3   7.0  11.6  25.5
```

```
370          RAJASTHAN         BARMER   1.2   1.7   2.3   3.2   9.1  27.4

      JUL   AUG   SEP   OCT   NOV   DEC  ANNUAL  Jan-Feb  Mar-May  Jun-Sep  \
353  13.1  14.1   8.6   7.2   3.6   4.0    94.6     15.4     24.8     39.6
374  59.3  56.9  22.8   3.1   1.5   1.0   181.2      4.6     12.6    158.4
359  11.6  24.0   8.6  14.5  20.6  20.5   223.3     51.2     66.8     49.7
373  79.3  64.9  31.7   3.1   2.1   4.4   252.9     14.0     27.9    201.4
370  88.6  86.7  40.7   4.1   3.2   0.4   268.6      2.9     14.6    243.4

     Oct-Dec
353     14.8
374      5.6
359     55.6
373      9.6
370      7.7
```

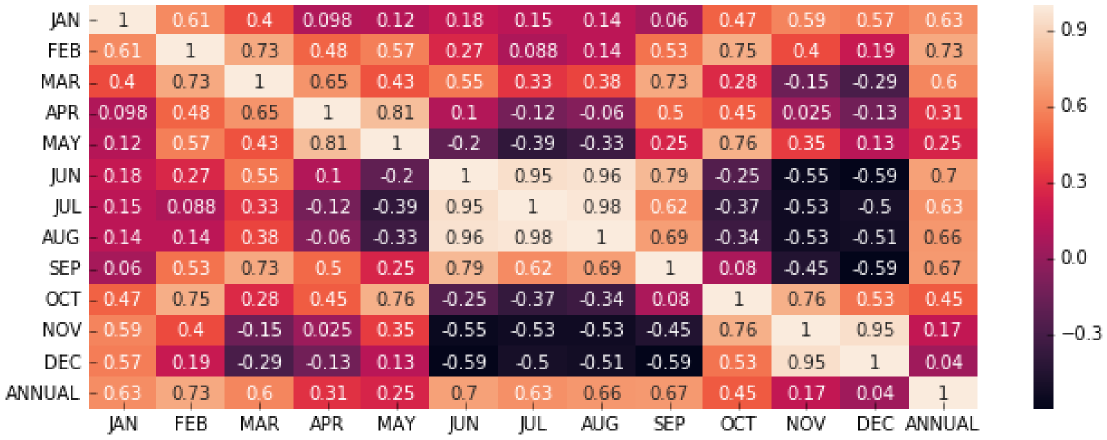
```
In [26]: fig = plt.figure(figsize=(15, 10))
         ax = plt.xticks(rotation=90)
         ax = sns.barplot(x='STATE_UT_NAME', y='ANNUAL', data=S)
         plt.ylabel('State/UT')
         plt.xlabel('Annual rainfall in mm')
         plt.title('Annual rainfall in different states of India')

Out[26]: <matplotlib.text.Text at 0x7efceb24b8d0>
```



```
In [27]: plt.figure(figsize=(11,4))
         sns.heatmap(ap_data[['Jan-Feb','Mar-May','Jun-Sep','Oct-Dec','ANNUAL']].corr(),annot=True)
         plt.show()
```



```
In [28]: plt.figure(figsize=(11,4))
         sns.heatmap(ap_data[['JAN','FEB','MAR','APR','MAY','JUN','JUL','AUG','SEP','OCT','NOV','DEC','
         plt.show()
```


JAN: 1 0.61 0.4 0.098 0.12 0.18 0.15 0.14 0.06 0.47 0.59 0.57 0.63
FEB: 0.61 1 0.73 0.48 0.57 0.27 0.088 0.14 0.53 0.75 0.4 0.19 0.73
MAR: 0.4 0.73 1 0.65 0.43 0.55 0.33 0.38 0.73 0.28 -0.15 -0.29 0.6
APR: 0.098 0.48 0.65 1 0.81 0.1 -0.12 -0.06 0.5 0.45 0.025 -0.13 0.31
MAY: 0.12 0.57 0.43 0.81 1 -0.2 -0.39 -0.33 0.25 0.76 0.35 0.13 0.25
JUN: 0.18 0.27 0.55 0.1 -0.2 1 0.95 0.96 0.79 -0.25 -0.55 -0.59 0.7
JUL: 0.15 0.088 0.33 -0.12 -0.39 0.95 1 0.98 0.62 -0.37 -0.53 -0.5 0.63
AUG: 0.14 0.14 0.38 -0.06 -0.33 0.96 0.98 1 0.69 -0.34 -0.53 -0.51 0.66
SEP: 0.06 0.53 0.73 0.5 0.25 0.79 0.62 0.69 1 0.08 -0.45 -0.59 0.67
OCT: 0.47 0.75 0.28 0.45 0.76 -0.25 -0.37 -0.34 0.08 1 0.76 0.53 0.45
NOV: 0.59 0.4 -0.15 0.025 0.35 -0.55 -0.53 -0.53 -0.45 0.76 1 0.95 0.17
DEC: 0.57 0.19 -0.29 -0.13 0.13 -0.59 -0.5 -0.51 -0.59 0.53 0.95 1 0.04
ANNUAL: 0.63 0.73 0.6 0.31 0.25 0.7 0.63 0.66 0.67 0.45 0.17 0.04 1
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
ANNUAL