Exploratory Data Analysis of Indian Rainfall Data

India is an agricultural country and secondary agro based market will be steady with a good monsoon. The economic growth of each year depends on the amount of duration of monsoon rain, bad monsoon can lead to destruction of some crops, which may result in scarcity of some agricultural products which in turn can cause food inflation, insecurity and public unrest. In our analysis we are trying to understand the behavior of rainfall in India over the years, by months and different

Describing the data:

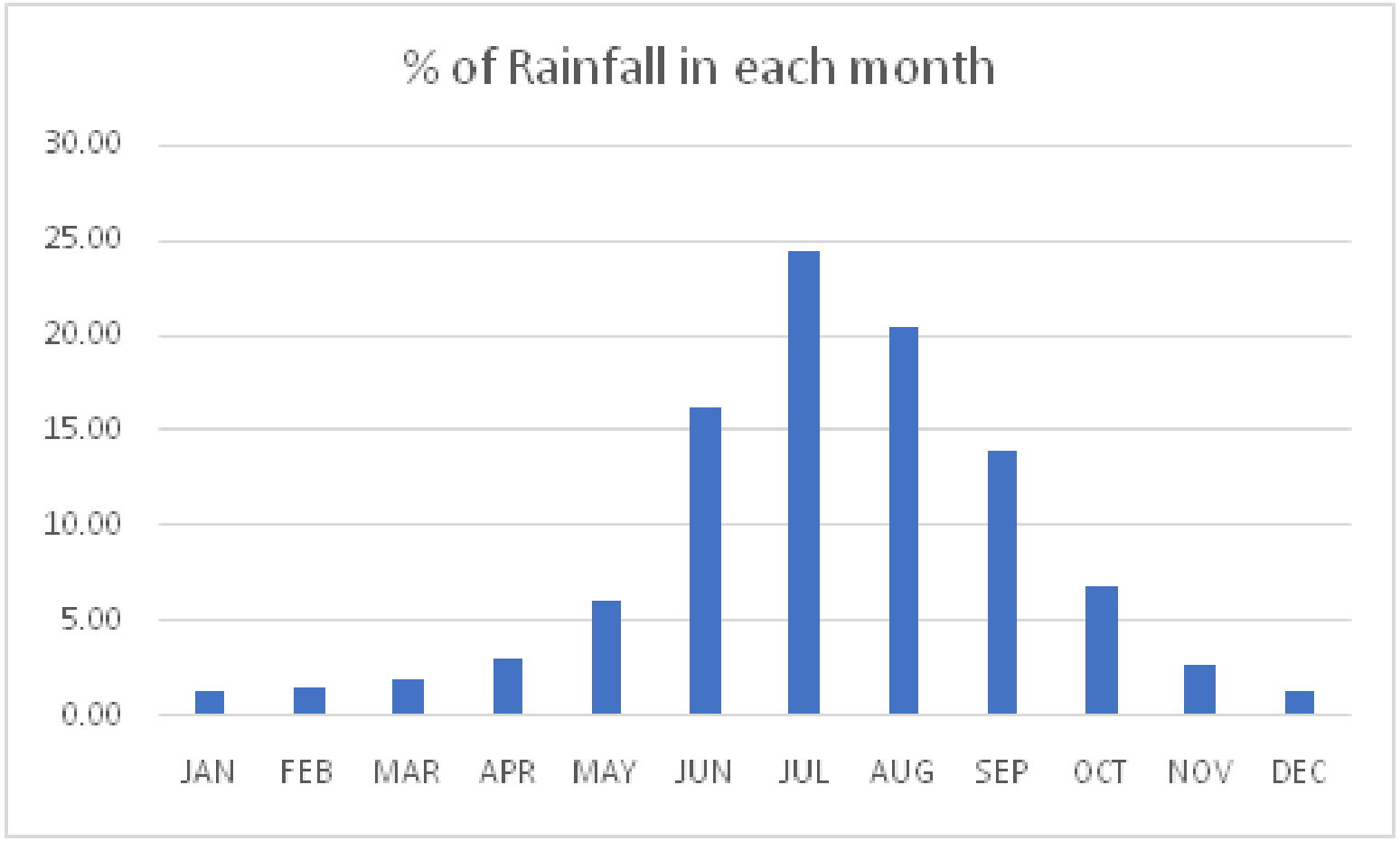
Data-set is downloaded from “data.gov.in” website. It has data for 117 years (1901–2017) consisting of monthly and seasonal data for all 36 meteorological subdivisions of India. So in total we have 117*12*36 = 50,544 observations. Our data-set had 0.7% of missing values. For the subdivision Arunachal Pradesh we had missing values for the first 15 years i.e. 1901 to 1915, so for all subdivisions we have considered data from 1916 to 2017 when we are analyzing as whole India. For the rest of the missing values we have used sequential imputation technique. Below table shows mean rainfall observed for each month over years. We can see that average rainfall is high in July and August followed by June and September.

Month	% of Rainfall
JAN	1.32
FEB	1.50
MAR	1.93
APR	3.06
MAY	6.13
JUN	16.25
JUL	24.45
AUG	20.42
SEP	13.99
OCT	6.81
NOV	2.79
DEC	1.34

Monthly mean annual rainfall

Annual rainfall by months:

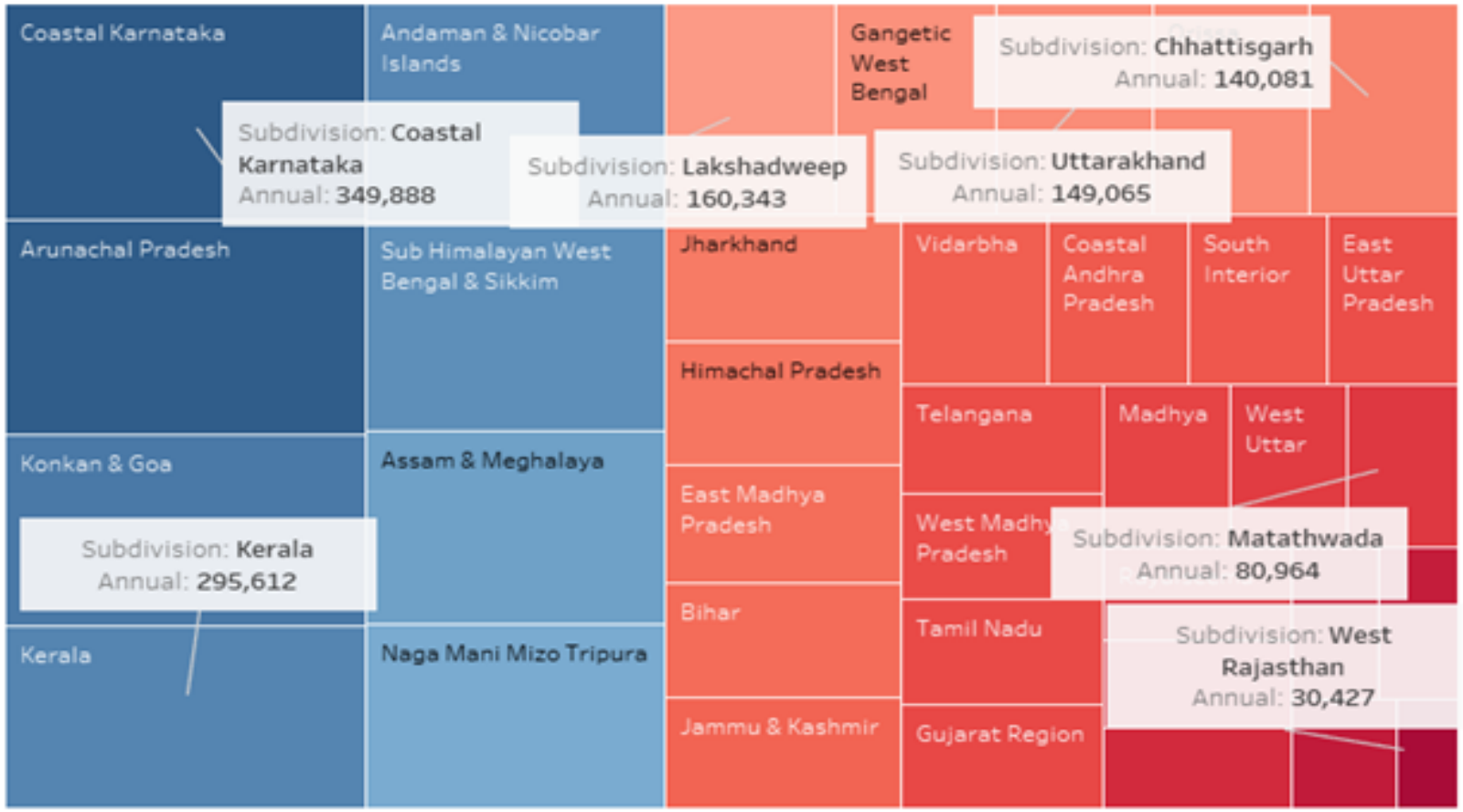
The below graph shows the percentage of rainfall each month receives when we consider India as a whole. The rainfall in the months of June, July, August and September together contribute to almost 80% of the annual rainfall.

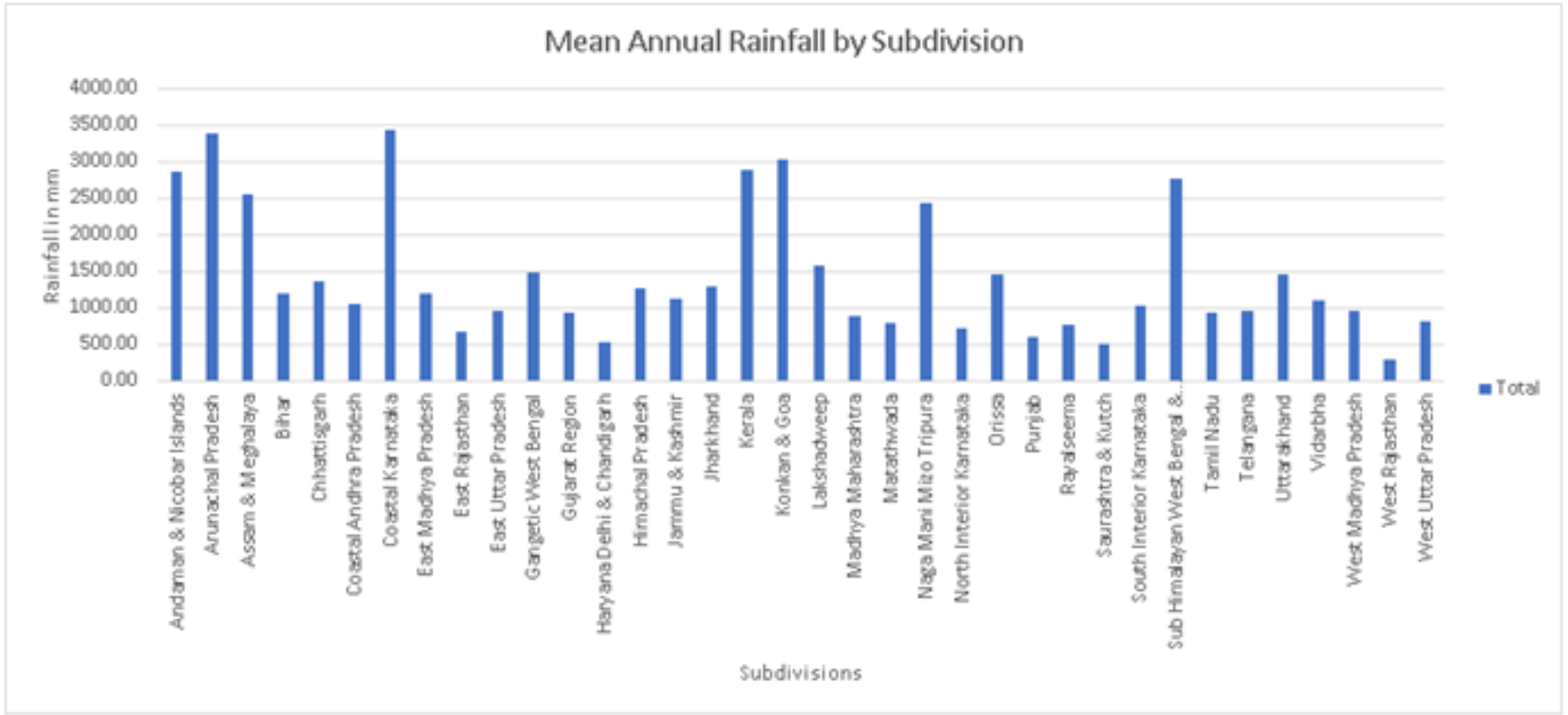


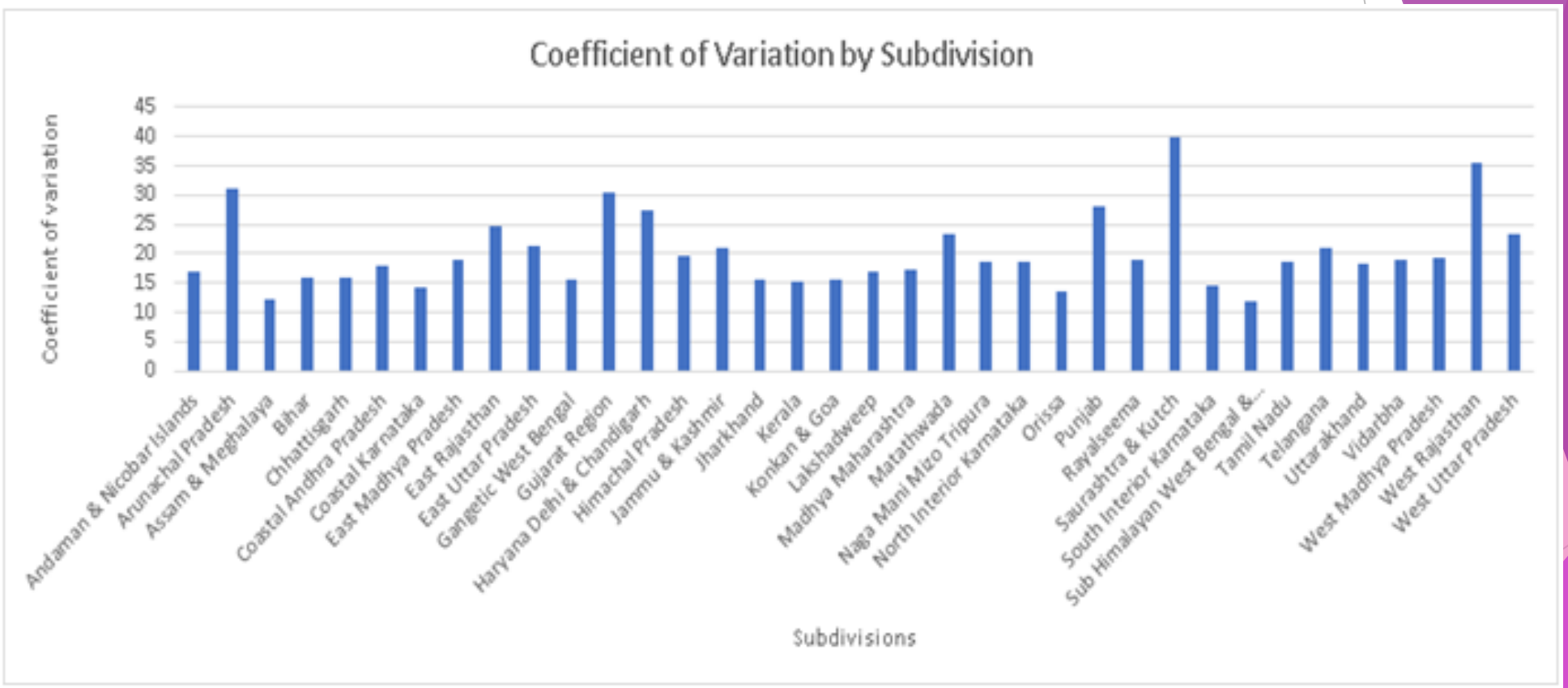
Annual rainfall by subdivision

The following is a heat map plotted based on sum of rainfall received by each subdivision for all these years. The subdivisions with large area represents high rainfall and with small boxes represent less rainfall. We can see that the subdivision located at Southwest and Northeast part of India have received more rainfall compared to central India.

Heat Map of Rainfall in all the Subdivisions




Mean Annual Rainfall by Subdivision
Rainfall in mm
4000.00
3500.00
3000.00
2500.00
2000.00
1500.00
1000.00
500.00
0.00
Andaman & Nicobar Islands
Arunachal Pradesh
Assam & Meghalaya
Bihar
Chhattisgarh
Coastal Andhra Pradesh
Coastal Karnataka
East Madhya Pradesh
East Rajasthan
East Uttar Pradesh
Gangetic West Bengal
Gujarat Region
Haryana Delhi & Chandigarh
Himachal Pradesh
Jammu & Kashmir
Jharkhand
Kerala
Konkan & Goa
Lakshadweep
Madhya Maharashtra
Matathwada
Naga Mani Mizo Tripura
North Interior Karnataka
Orissa
Punjab
Rayalseema
Saurashtra & Kutch
South Interior Karnataka
Sub Himalayan West Bengal &...
Tamil Nadu
Telangana
Uttarakhand
Vidarbha
West Madhya Pradesh
West Rajasthan
West Uttar Pradesh
Subdivisions
Total


Coefficient of Variation by Subdivision
Coefficient of variation
45
40
35
30
25
20
15
10
5
0
Andaman & Nicobar Islands
Arunachal Pradesh
Assam & Meghalaya
Bihar
Chhattisgarh
Coastal Andhra Pradesh
Coastal Karnataka
East Madhya Pradesh
East Rajasthan
East Uttar Pradesh
Gangetic West Bengal
Gujarat Region
Haryana Delhi & Chandigarh
Himachal Pradesh
Jammu & Kashmir
Jharkhand
Kerala
Konkan & Goa
Lakshadweep
Madhya Maharashtra
Matathwada
Naga Mani Mizo Tripura
North Interior Karnataka
Orissa
Punjab
Rayalseema
Saurashtra & Kutch
South Interior Karnataka
Sub Himalayan West Bengal &...
Tamil Nadu
Telangana
Uttarakhand
Vidarbha
West Madhya Pradesh
West Rajasthan
West Uttar Pradesh
Subdivisions

Understanding the Monsoon in India:

We will now move to a more interesting part of the analysis.

We will see what exactly is monsoon, different types of monsoon winds in India, which subdivisions of India receives rainfall from which monsoon winds and why only particular subdivisions receive highest rainfall during this monsoon season.

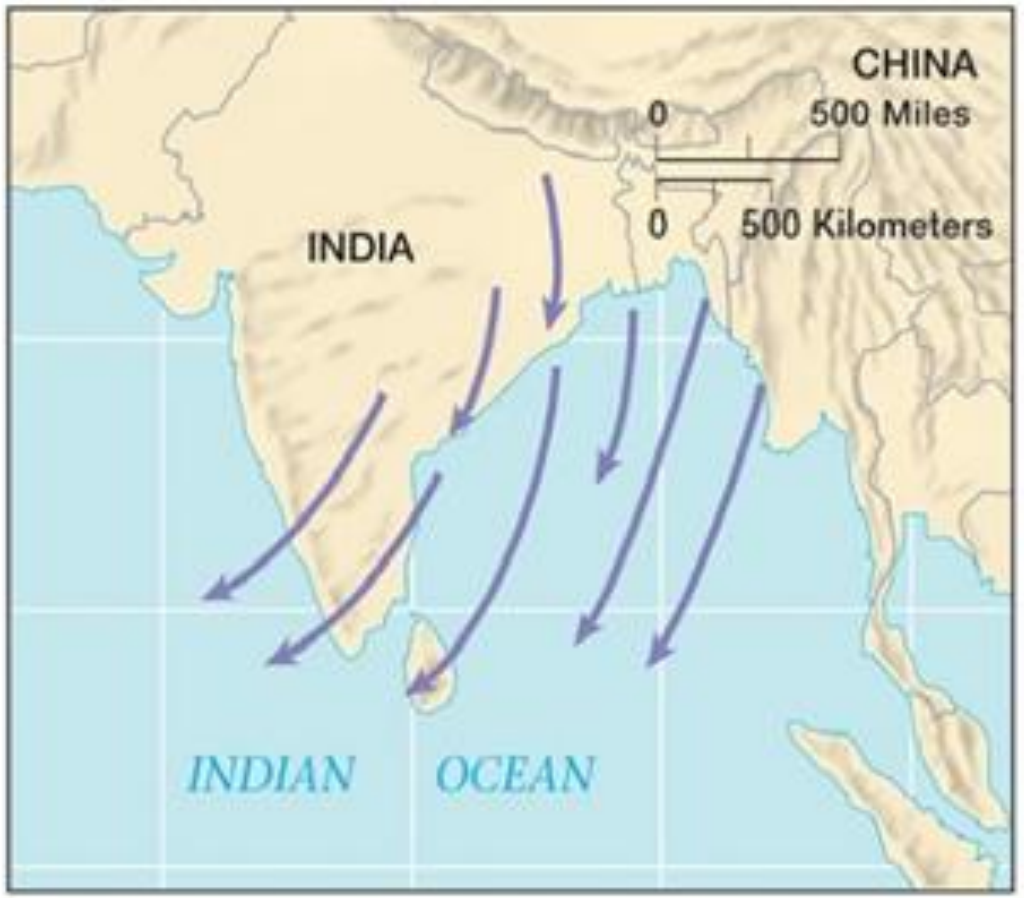
Before addressing these questions, we will see the basic concept of how rainfall actually occurs. So basically, during the summers, the Indian subcontinent heats up more as compared to the Indian ocean as the sun is directly over the landmass. This creates a low pressure over the Indian subcontinent and a relatively low pressure over the Indian ocean. And as we know, the wind flows from high pressure area to the low pressure are in order to fill the void that was created thanks to the pressure system. So when the wind starts flowing from high pressure area to low pressure are i.e. from sea to land, it picks up the moisture from the sea and while entering the indian subcontinent it comes in contact with the high terrains and hence precipitation occurs.

Now that we have understood how rainfall occurs we will now see what exactly is monsoon? Monsoon is traditionally defined as a seasonal reversal of wind accompanied by corresponding changes in precipitation.

As seen in the image below, India receives rainfall fromSouthwest Monsoon winds (Summer Monsoon or Advancing Monsoon) and Northeast Monsoon winds (Winter Monsoon or Retreating Monsoon)

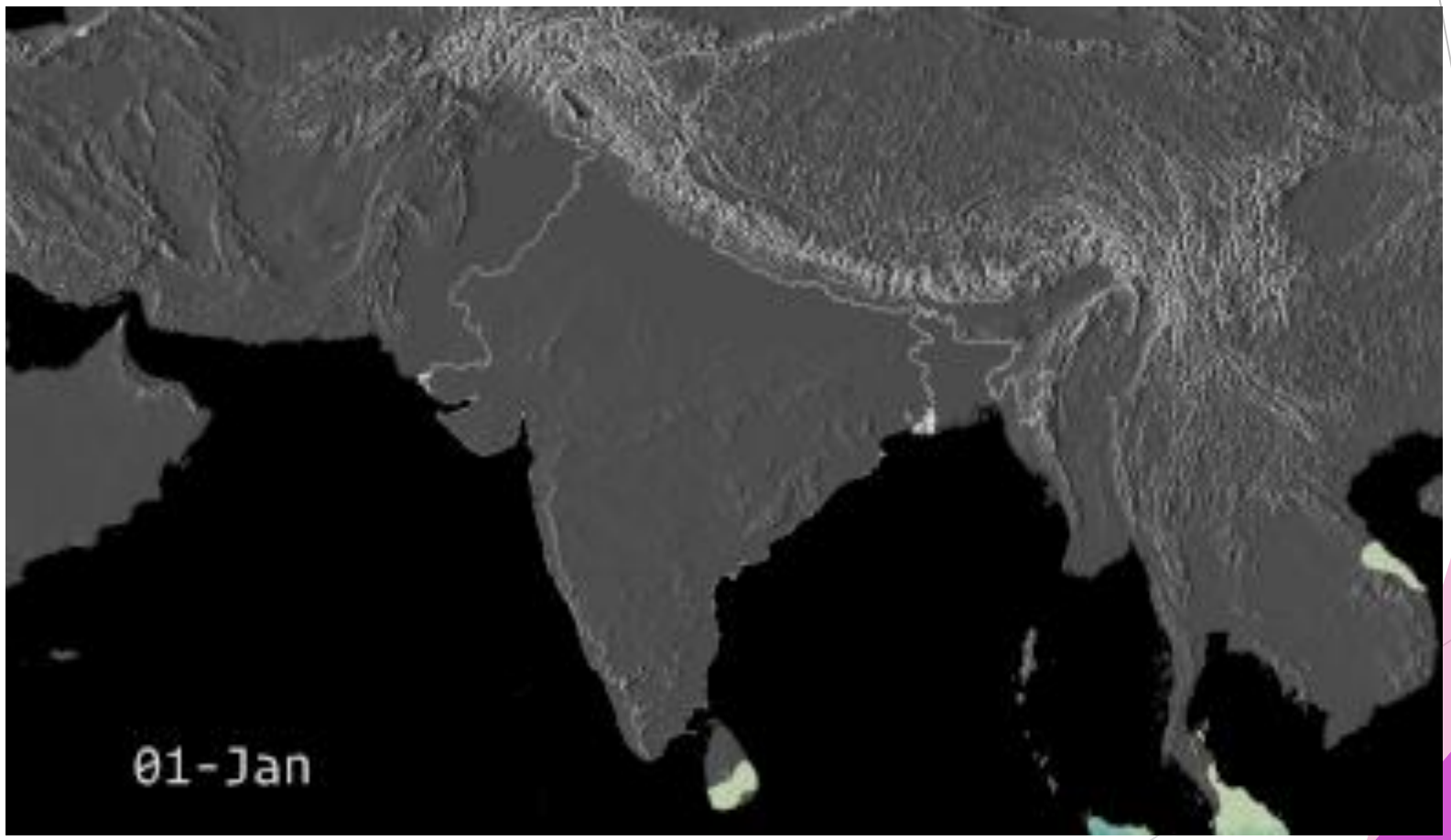
The Southwest Monsoon usually starts in the first week of June and ends by first week of September and monsoon usually starts retreating from the Indian Subcontinent by the start of September and leaves the subcontinent completely by the end of November. And as we have seen in the previous graphs that Southwest monsoon provides almost 80% of the rainfall in India. This Southwest Monsoon has two branches, namely Arabian Sea branch and

Summer

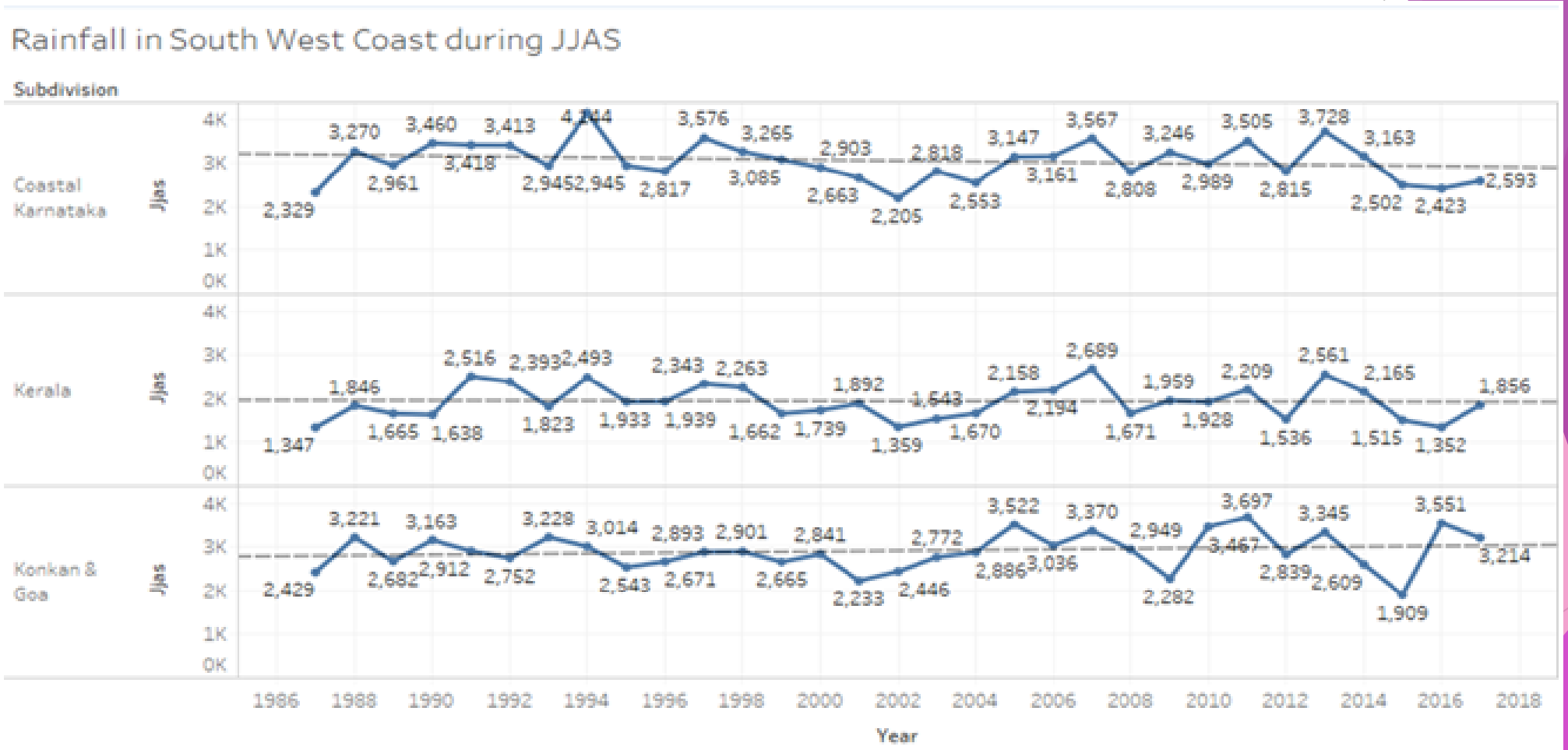


Winter

The following figure shows the cycle of monsoon in India over a year

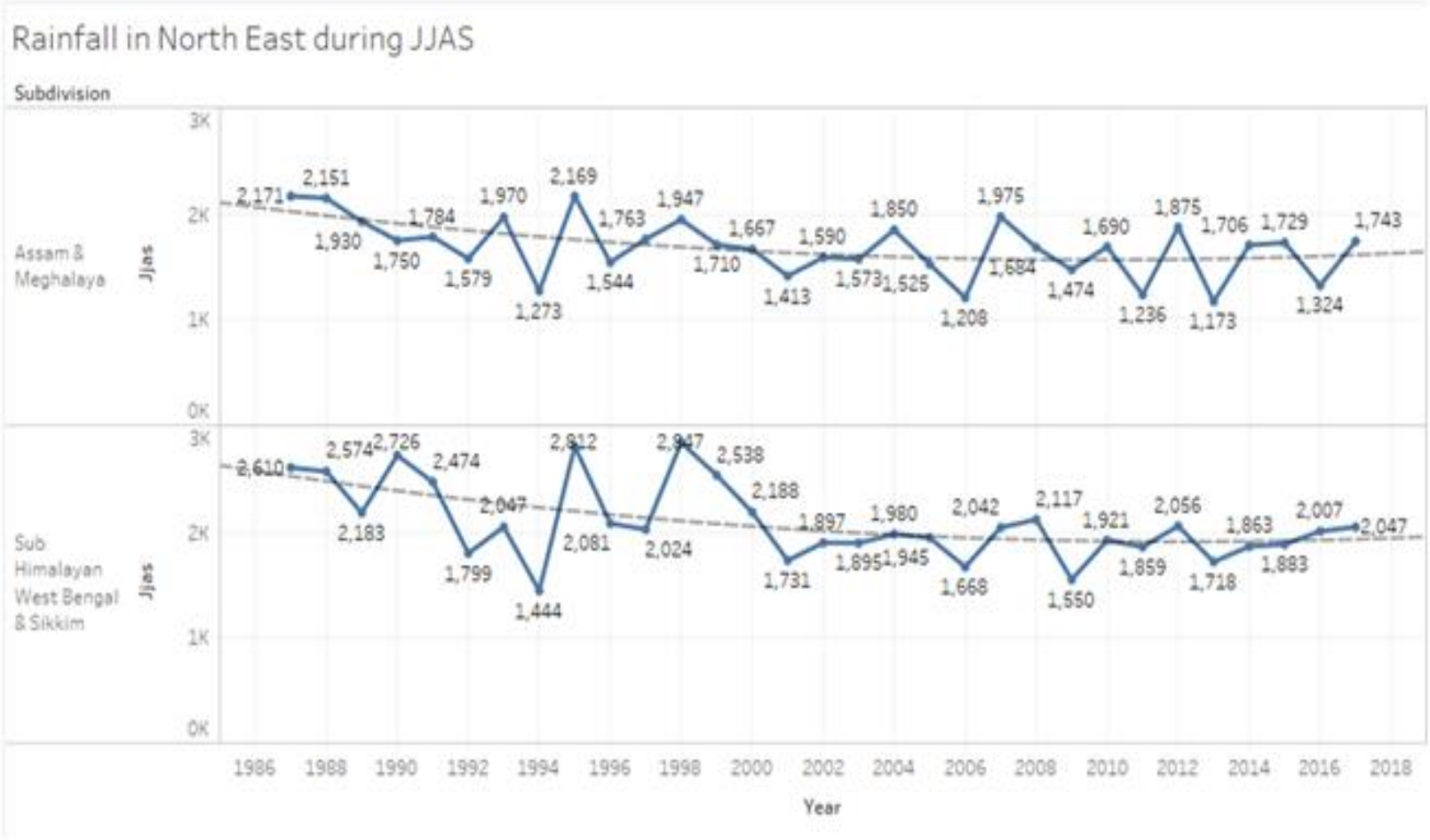


Now, due to the presence of high rising Western Ghats which runs along the South West coast of India in the states of Kerala, Karnataka, Goa and Maharashtra, they block the Arabian Sea branch of southwest monsoon and hence these regions receive very high rainfall during monsoon season. This is shown in the graph below. (For all the graphs we have considered the last 3 decades i.e. last 30 years data (1987–2017))


Rainfall in South West Coast during JJAS
Subdivision
Coastal Karnataka
Kerala
Konkan & Goa
Jjas
4K
3K
2K
1K
0K
2,329 3,270 2,961 3,460 3,418 3,413 2,945 4,244 2,945 2,817 3,576 3,085 3,265 2,663 2,903 2,205 2,818 2,553 3,147 3,161 3,567 2,808 3,246 2,989 3,505 2,815 3,728 2,502 3,163 2,423 2,593
1,347 1,846 1,665 1,638 2,516 1,823 2,393 2,493 1,933 2,343 1,939 2,263 1,662 1,739 1,892 1,359 1,643 1,670 2,158 2,194 2,689 1,671 1,959 1,928 2,209 1,536 2,561 1,515 2,165 1,352 1,856
2,429 3,221 2,682 3,163 2,912 2,752 3,228 3,014 2,543 2,893 2,671 2,901 2,665 2,841 2,233 2,772 2,446 3,522 2,886 3,370 3,036 2,949 2,282 3,697 3,467 3,345 2,839 2,609 1,909 3,551 3,214
1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2018
Year

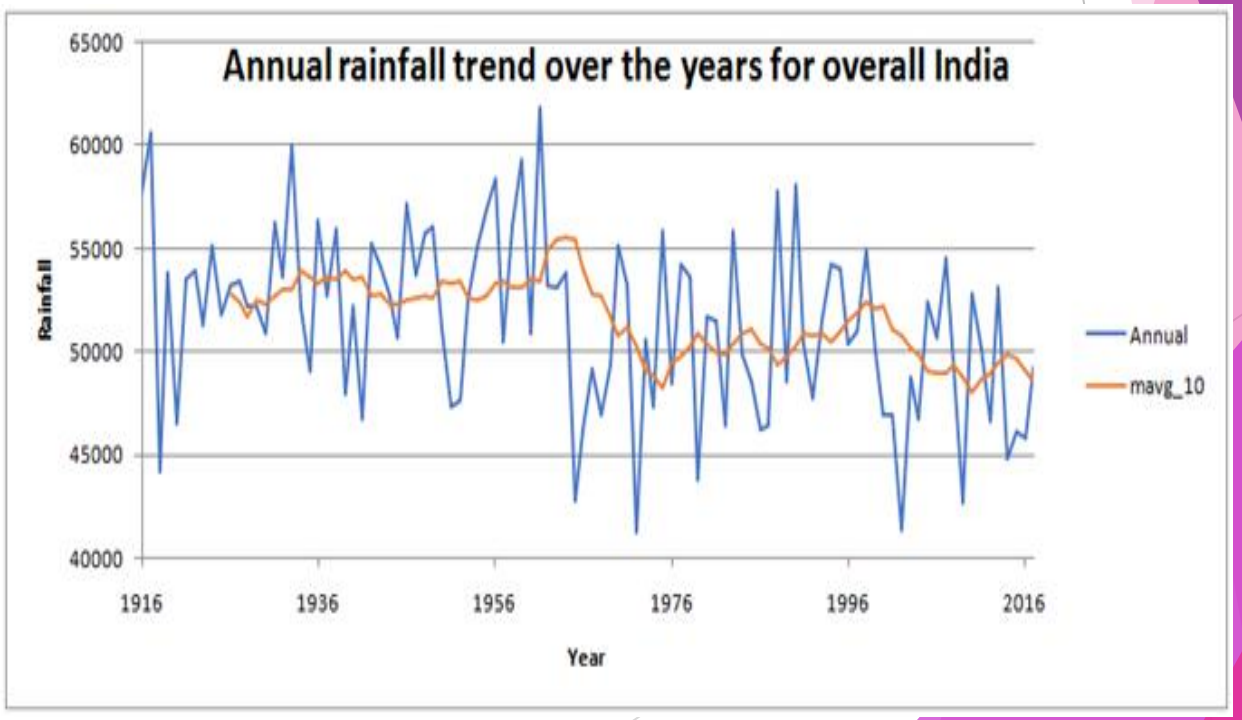
The Bay of Bengal branch of the Southwest Monsoon provides rainfall to the eastern and north eastern regions of India. Due to the presence of the eastern ghats along Orissa and West Bengal, these regions receive high rainfall in the East and when this branch reaches the north eatsern part of India i.e. Sikkim, Arunachal Pradesh, Assam & Meghalaya, due to the presence of Khasi hills and other hilly areas, these subdivisions receive very high rainfall. Cherrapunji in Meghalaya is titled as one of the wettest places on earth.

The average annual rainfall in these regions is shown in the graphs below.



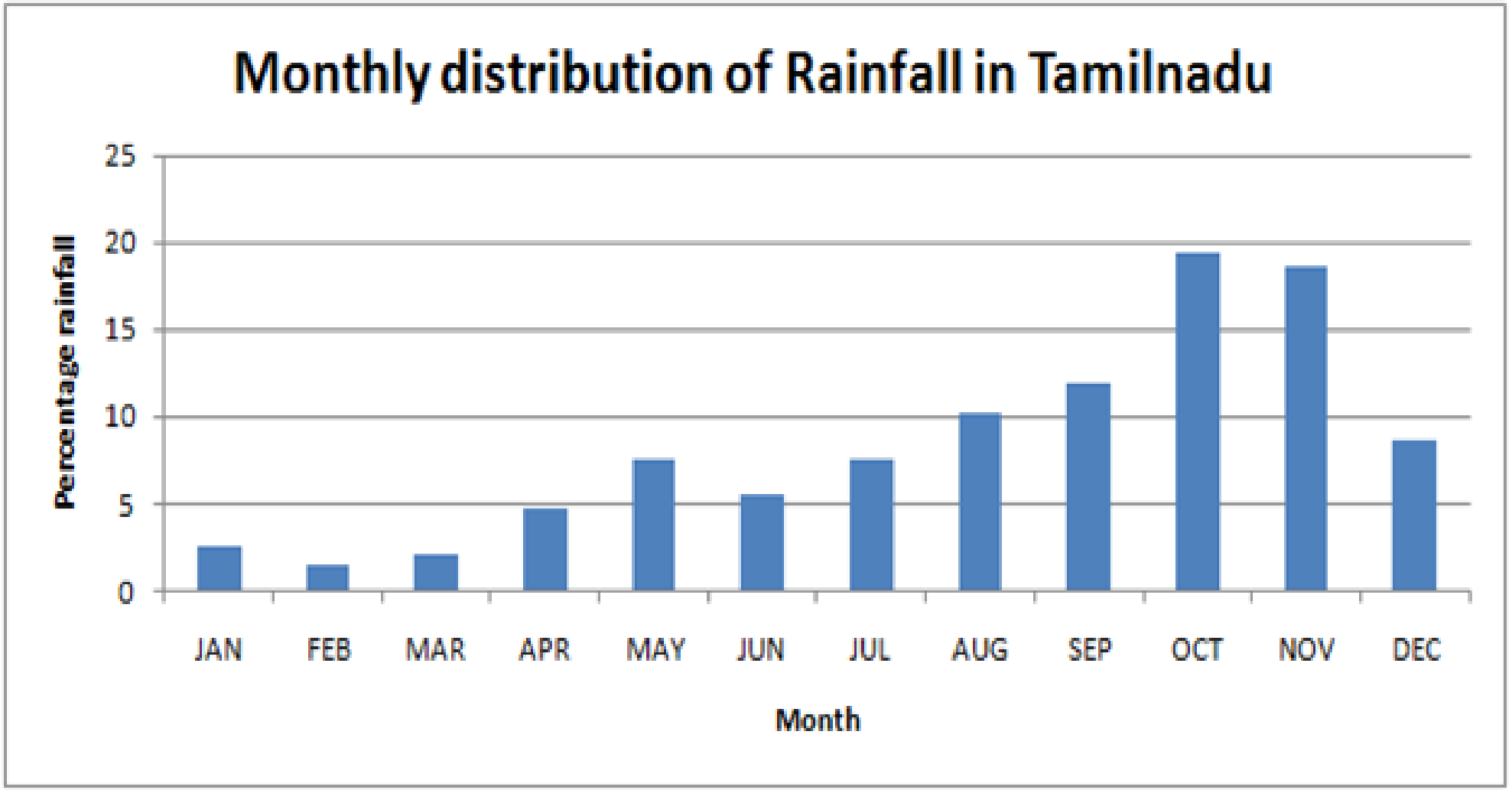
Annual Rainfall trend over the years for whole India

10 years moving average was plotted, we can see that there is a decreasing trend in rainfall in the recent years.



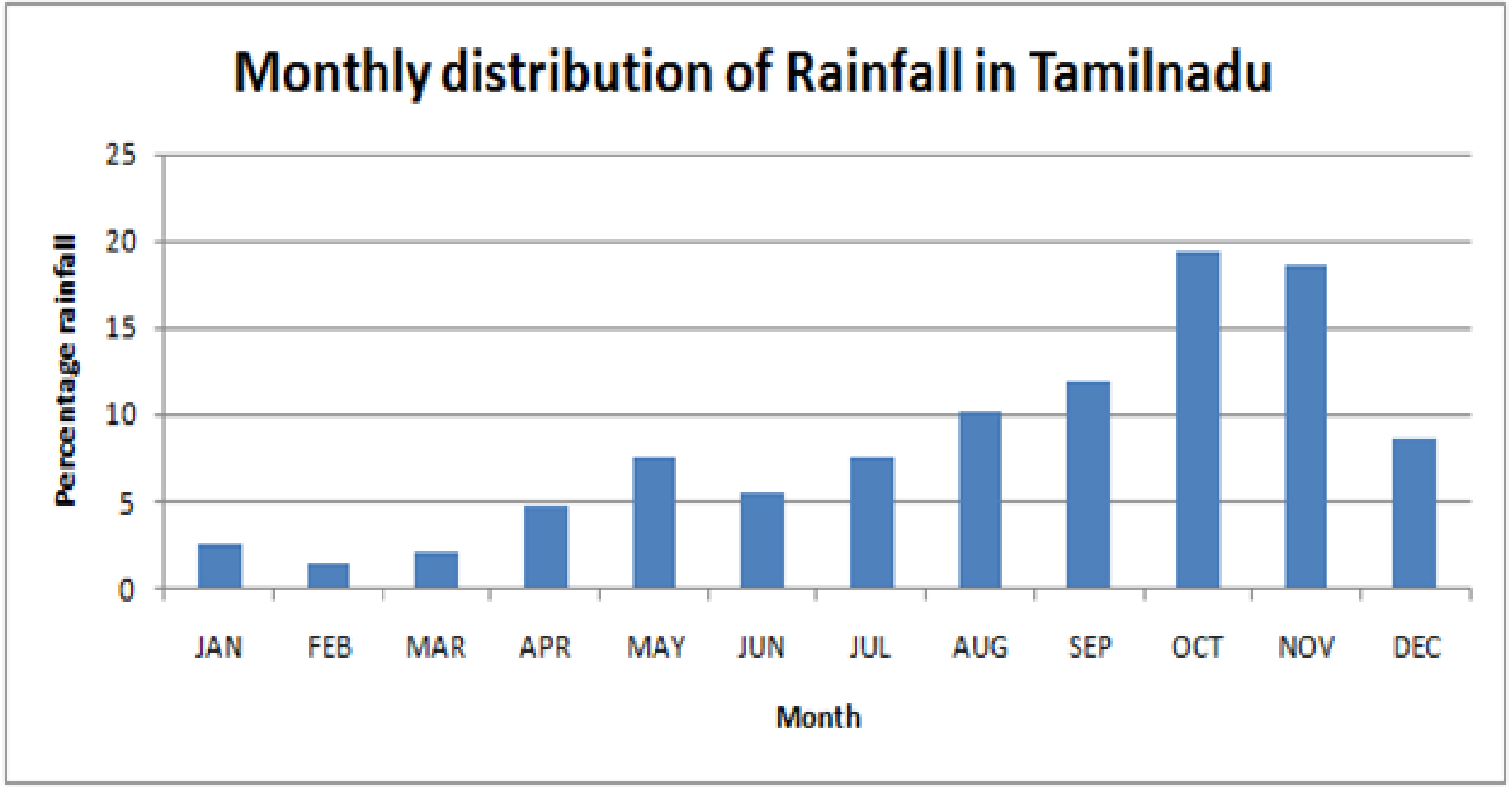
Rainfall Data Analysis for Subdivision TamilNadu

To do rainfall data analysis of Tamil Nadu subdivision, we have considered the data from year 1901 to 2017. First we will see the distribution of rainfall over months and we can see in below figure that rainfall is more in October and November compared to other months as Tamilnadu receives rain during retreating monsoon season because of Northeast trade winds. The climate of Tamil Nadu is tropical in nature with less variation in temperature in summer and winter. This is because of its geographical location.


Monthly distribution of Rainfall in Tamilnadu
Percentage rainfall
25
20
15
10
5
0
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
Month

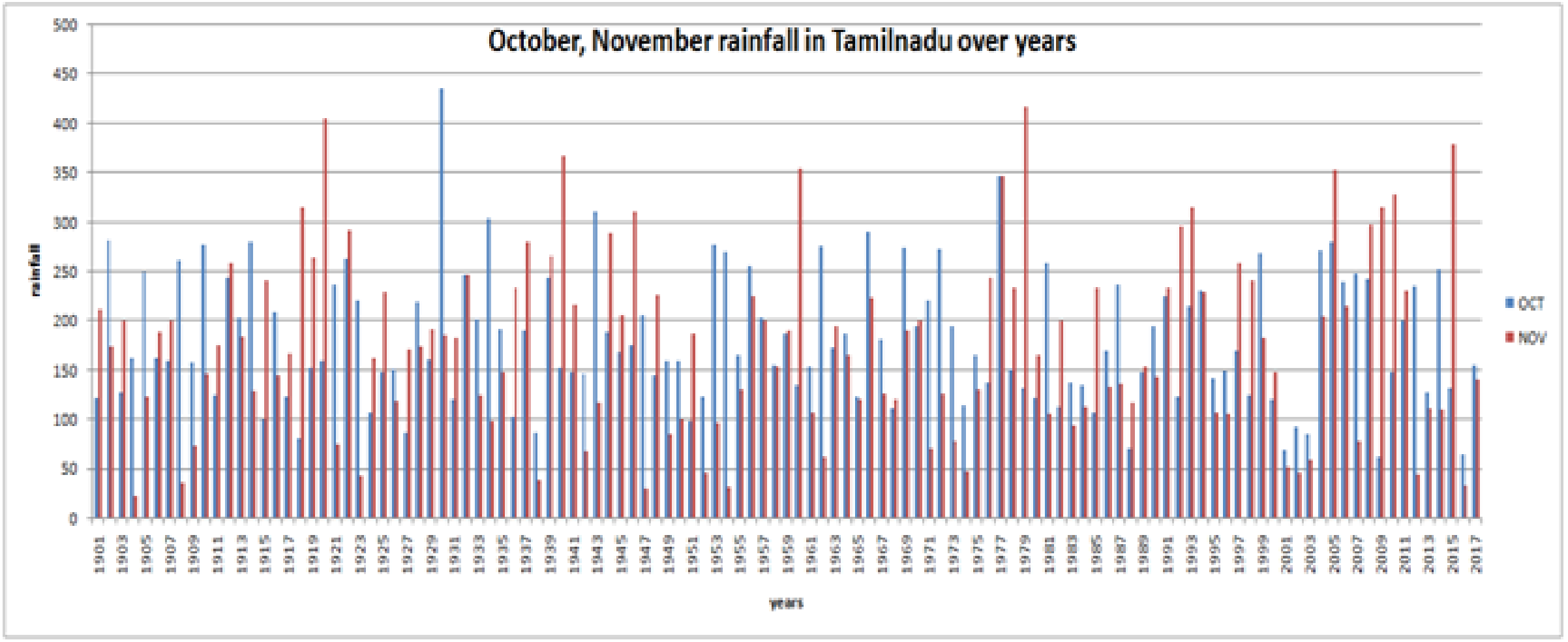
Rainfall Data Analysis for Subdivision TamilNadu

To do rainfall data analysis of Tamil Nadu subdivision, we have considered the data from year 1901 to 2017. First we will see the distribution of rainfall over months and we can see in below figure that rainfall is more in October and November compared to other months as Tamilnadu receives rain during retreating monsoon season because of Northeast trade winds. The climate of Tamil Nadu is tropical in nature with less variation in temperature in summer and winter. This is because of its geographical location.


Monthly distribution of Rainfall in Tamilnadu
Percentage rainfall
25
20
15
10
5
0
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
Month

The below graph shows rainfall in Tamil Nadu in months October and November here we can see that most of the years when there is very low rainfall in October there is very high rainfall in November and vice versa. Tamil Nadu also rainfalls from tropical cyclones emerging in the neighborhood of Andaman islands during the retreat monsoon.

In 2015 there were south Indian floods, the flooding in Chennai was described as the worst in a century. In winters of 2017 the worst drought in a century happened in southern India.


October, November rainfall in Tamilnadu over years
rainfall
years
OCT
NOV
0
50
100
150
200
250
300
350
400
450
500
1901
1903
1905
1907
1909
1911
1913
1915
1917
1919
1921
1923
1925
1927
1929
1931
1933
1935
1937
1939
1941
1943
1945
1947
1949
1951
1953
1955
1957
1959
1961
1963
1965
1967
1969
1971
1973
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
2015
2017

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib. Pyp lot as plt
import seaborn as sns
import random
# Input data files are available in the "../input/" directory.
# For example, running this (by clicking run or pressing Shift+Enter)
will list all files under the input directory
import os
for dirname, _, filenames in os. walk('/kaggle/input'):
    for filename in filenames:
        print(os .path. join(dirname, filename))
```

In [7]:

```
data = pd. read csv('/content/rainfall_India_2017
(1).csv').rename(columns=str. lower)
data. head(3)
```

Out[7]:

	subdivision	year	jan	feb	mar	apr	may	jun		aug	sep	oct	nov	dec	annual
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	558.2	33.6	3373.2
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	359.0	160.5	3520.7
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	284.4	225.0	2957.4

```
data.info()
RangeIndex: 4188 entries, 0 to 4187
Data columns (total 15 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   subdivision  4188 non-null   object 
 1   year         4188 non-null   int64  
 2   jan          4184 non-null   float64
 3   feb          4185 non-null   float64
 4   mar          4182 non-null   float64
 5   apr          4184 non-null   float64
 6   may          4185 non-null   float64
 7   jun          4183 non-null   float64
 8   jul          4181 non-null   float64
 9   aug          4184 non-null   float64
 10  sep          4182 non-null   float64
 11  oct          4181 non-null   float64
 12  nov          4177 non-null   float64
 13  dec          4178 non-null   float64
 14  annual       4162 non-null   float64
dtypes: float64(13), int64(1), object(1)
memory usage: 490.9+ KB
```

```
print('Dataset comprises of {} observations and {}
characteristics'.format(data.shape[0],data.
shape[1]))
print('\nUnique Values: ',data.nunique())
print('\nMissing Values:
',data.isna().sum())
```

```
Dataset comprises of 4188 observations and
15 characteristics

Unique Values:  subdivision      36
year             117
jan              808
feb              902
mar              989
apr             1247
may             1751
jun             2754
jul             3093
aug             2950
sep             2664
oct             1958
nov             1245
dec              810
annual          3770
dtype: int64
```

```
Missing Values:  subdivision     0
year              0
jan               4
feb               3
mar               6
apr               4
may               3
jun               5
jul               7
aug               4
sep               6
oct               7
nov              11
dec              10
annual           26
dtype: int64
```

raw.githubusercontent.com/IBM-EPBL/IBM-Project-25888-1659976192/main/Development%20phase/Sprint1/Capture.JPG

Type here to search

27°C ENG 11:11 14-11-2022

THANKING YOU

THANKING YOU