

DATA VISUALIZATION DIGITAL ASSIGNMENT 2

REPORT

SLETZER CONCY MASCARENHAS

20BCE1439

D2 SLOT

TOPIC : Analysis Of Rainfall in India

Dataset Description

- Dataset 1(rainfall_in_India 901-2015):

This dataset Consists of 19 columns and 4117 rows .The first column contains the State name,the second column contains the year of the recorded rainfall,the next 12 columns contain the rainfall received in mm of each month of the year,the next column contains the annual rainfall received during the year and the final 3 columns contains the aggregate rainfall received during the 4 seasonal months respectively.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	SUBDIVISI	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec
2	ANDAMAN	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	136.3	560.3	1696.3	980.3
3	ANDAMAN	1902	0	159.8	12.2	0	446.1	537.1	228.9	753.7	666.2	197.2	359	160.5	3520.7	159.8	458.3	2185.9	716.7
4	ANDAMAN	1903	12.7	144	0	1	235.1	479.9	728.4	326.7	339	181.2	284.4	225	2957.4	156.7	236.1	1874	690.6
5	ANDAMAN	1904	9.4	14.7	0	202.4	304.5	495.1	502	160.1	820.4	222.2	308.7	40.1	3079.6	24.1	506.9	1977.6	571
6	ANDAMAN	1905	1.3	0	3.3	26.9	279.5	628.7	368.7	330.5	297	260.7	25.4	344.7	2566.7	1.3	309.7	1624.9	630.8
7	ANDAMAN	1906	36.6	0	0	0	556.1	733.3	247.7	320.5	164.3	267.8	128.9	79.2	2534.4	36.6	556.1	1465.8	475.9
8	ANDAMAN	1907	110.7	0	113.3	21.6	616.3	305.2	443.9	377.6	200.4	264.4	648.9	245.6	3347.9	110.7	751.2	1327.1	1158.9
9	ANDAMAN	1908	20.9	85.1	0	29	562	693.6	481.4	699.9	428.8	170.7	208.1	196.9	3576.4	106	591	2303.7	575.7
10	ANDAMAN	1910	26.6	22.7	206.3	89.3	224.5	472.7	264.3	337.4	626.6	208.2	267.3	153.5	2899.4	49.3	520.1	1701	629
11	ANDAMAN	1911	0	8.4	0	122.5	327.3	649	253	187.1	464.5	333.8	94.5	247.1	2687.2	8.4	449.8	1553.6	675.4
12	ANDAMAN	1912	583.7	0.8	0	21.9	140.7	549.8	468.9	370.3	386.2	318.7	117.2	2.3	2960.5	584.5	162.6	1775.2	438.2
13	ANDAMAN	1913	84.8	0.5	1.3	2.5	190.7	530	280.8	205.8	580.1	288.8	133	67.5	2365.8	85.3	194.5	1596.7	489.3
14	ANDAMAN	1914	0	0	0	37.7	298.8	383.3	792.8	520.5	310.8	139.8	184.4	289.7	2957.8	0	336.5	2007.4	613.9
15	ANDAMAN	1915	45	56.7	33.3	40.9	170.2	334.7	269	317.2	429.8	468.1	258.4	318	2741.3	101.7	244.4	1350.7	1044.5
16	ANDAMAN	1916	0	0	0	0.5	487.4	450.1	317.3	425	561.2	369.7	192.6	133.7	2937.5	0	487.9	1753.6	696

- Dataset 2(district wise rainfall):

This dataset consists of 19 columns and 642 rows.The first row gives us the state name,the second row contains a single district belonging to the state,the next 12 columns contain the rainfall received during each of the 12 months of the year in mm,the next column contains the annual rainfall in mm received by the district,the next 4 columns contain the rainfall received by the districts based on the seasonal months respectively.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	STATE_UT	DISTRICT	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec
2	ANDAMAN	NICOBAR	107.3	57.9	65.2	117	358.5	295.5	285	271.9	354.8	326	315.2	250.9	2805.2	165.2	540.7	1207.2	892.1
3	ANDAMAN	SOUTH ANDA	43.7	26	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	705.3
4	ANDAMAN	N & M ANDA	32.7	15.9	8.6	53.4	343.6	503.3	465.4	460.9	454.8	276.1	198.6	100	2913.3	48.6	405.6	1884.4	574.7
5	ARUNACH	LOHIT	42.2	80.8	176.4	358.5	306.4	447	660.1	427.8	313.6	167.1	34.1	29.8	3043.8	123	841.3	1848.5	231
6	ARUNACH	EAST SIANG	33.3	79.5	105.9	216.5	323	738.3	990.9	711.2	568	206.9	29.5	31.7	4034.7	112.8	645.4	3008.4	268.1
7	ARUNACH	SUBANSIRI F.	28	48.3	85.3	101.5	140.5	228.4	217.4	182.8	159.8	75.9	20.9	11.6	1300.4	76.3	327.3	788.4	108.4
8	ARUNACH	TIRAP	42.2	72.7	141	316.9	328.7	614.7	851.9	500.6	418.3	218.7	42.9	22.9	3571.5	114.9	786.6	2385.5	284.5
9	ARUNACH	ANJAW (LOH	42.2	80.8	176.4	358.5	306.4	447	660.1	427.8	313.6	167.1	34.1	29.8	3043.8	123	841.3	1848.5	231
10	ARUNACH	LOWER DIBA	83.7	153.9	303.5	383.6	268	374.2	272	160.5	266.7	167.2	64	56	2553.3	237.6	955.1	1073.4	287.2
11	ARUNACH	CHANGLANG	70.3	170.9	367.9	554.4	334.2	526.2	460.8	291.5	353.6	275	64.9	74.2	3543.9	241.2	1256.5	1632.1	414.1
12	ARUNACH	PAPUM PARE	33.5	67.8	106.1	226.9	453	640.5	609.5	503.4	492.3	214.7	19.2	11.3	3378.2	101.3	786	2245.7	245.2
13	ARUNACH	LOW SUBAN	97.5	109.3	92.4	204.3	266.2	284.1	248.9	270.5	192.7	78.5	49.5	27.2	1921.1	206.8	562.9	996.2	155.2
14	ARUNACH	UPPER SIANG	74.3	176.7	362.6	397.5	408.7	801.9	653	417.9	686	264.9	86.9	71.7	4402.1	251	1168.8	2558.8	423.5
15	ARUNACH	WEST SIANG	26	66.7	76.8	229.2	239.5	416.6	592.4	312.4	291.1	126.8	33.7	29.5	2440.7	92.7	545.5	1612.5	190
16	ARUNACH	DIBANG VALL	83.7	153.9	303.5	383.6	268	374.2	272	160.5	266.7	167.2	64	56	2553.3	237.6	955.1	1073.4	287.2
17	ARUNACH	WEST KAMEN	35.2	43.5	58.9	134.3	341.1	665.3	749.9	579.1	490.9	233.9	40.3	27	3399.4	78.7	534.3	2485.2	301.2
18	ARUNACH	EAST KAMEN	49	74.4	96.5	156.9	208	345.7	368.5	256.2	275.9	138.2	34.4	27.2	2030.9	123.4	461.4	1246.3	199.8
19	ARUNACH	TAWANG(W I	35.2	43.5	58.9	134.3	341.1	665.3	749.9	579.1	490.9	233.9	40.3	27	3399.4	78.7	534.3	2485.2	301.2
20	ARUNACH	KURUNG KUI	82.7	70	128.2	245.7	271.4	292.7	404	276.3	283.5	92.3	32.3	42.4	2221.5	152.7	645.3	1256.5	167
21	ASSAM	CACHAR	13.3	50.2	168.3	262.5	386.4	532.1	526.2	470.8	360.8	182.4	34.8	11.4	2999.2	63.5	817.2	1889.9	228.6
22	ASSAM	DARRANG	13.1	21.4	53.5	168.8	320	419.7	345.8	272.1	221.5	95.4	17.2	9.3	1957.8	34.5	542.3	1259.1	121.9

Dataset Link:

<https://www.kaggle.com/datasets/rajanand/rainfall-in-india/code>

Data Preprocessing

Basic preprocessing function have been carried out using Inbuilt functions to ignore NA values while performing visualizations and calculative operations

Dataset Join creation

Basic direct join is done between both the datasets where the Subdivision name is equivalent to the State/UT Name

district wise rainfall norm...

rainfall in india 1901-2015....

district wi... — rainfall in i...

How do relationships differ from joins? [Learn more](#)

district wise rainfall ...	Operator	rainfall in india 1901...
Abc State Ut Name	=	Abc Subdivision

+ Add more fields

> Performance Options

district wise rainfall normal.csv

State Ut Name

ANDAMAN And NICOBAR IS

ANDAMAN And NICOBAR IS

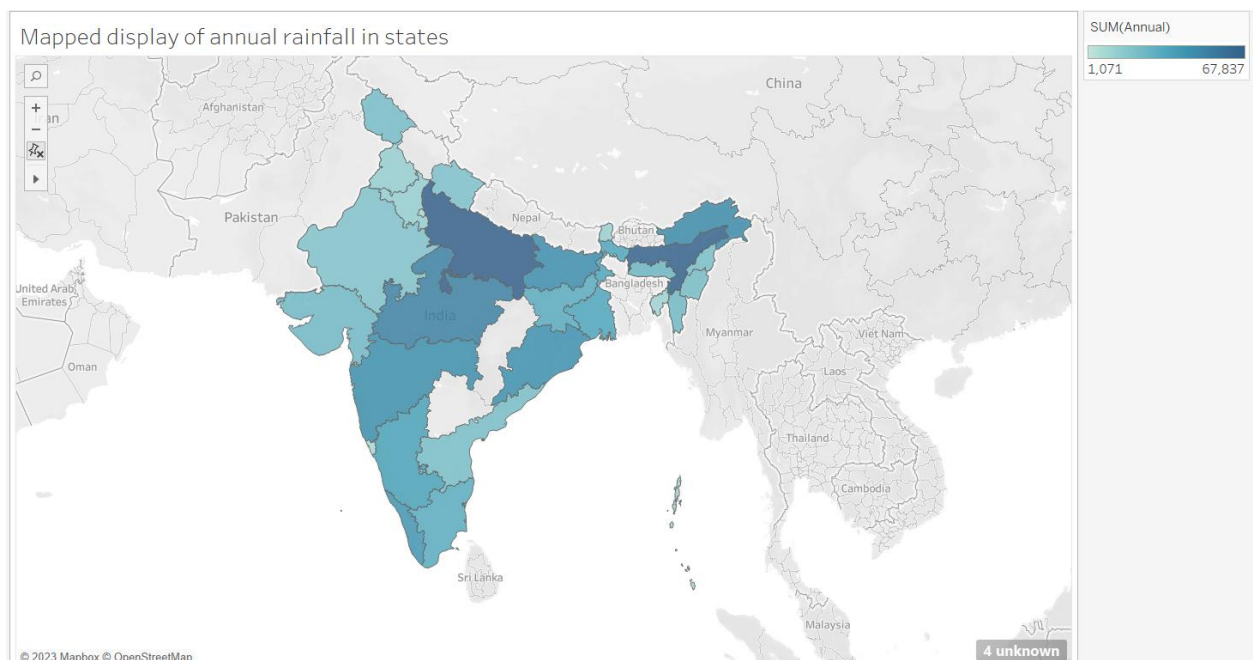
ANDAMAN And NICOBAR IS

ARUNACHAL PRADESH

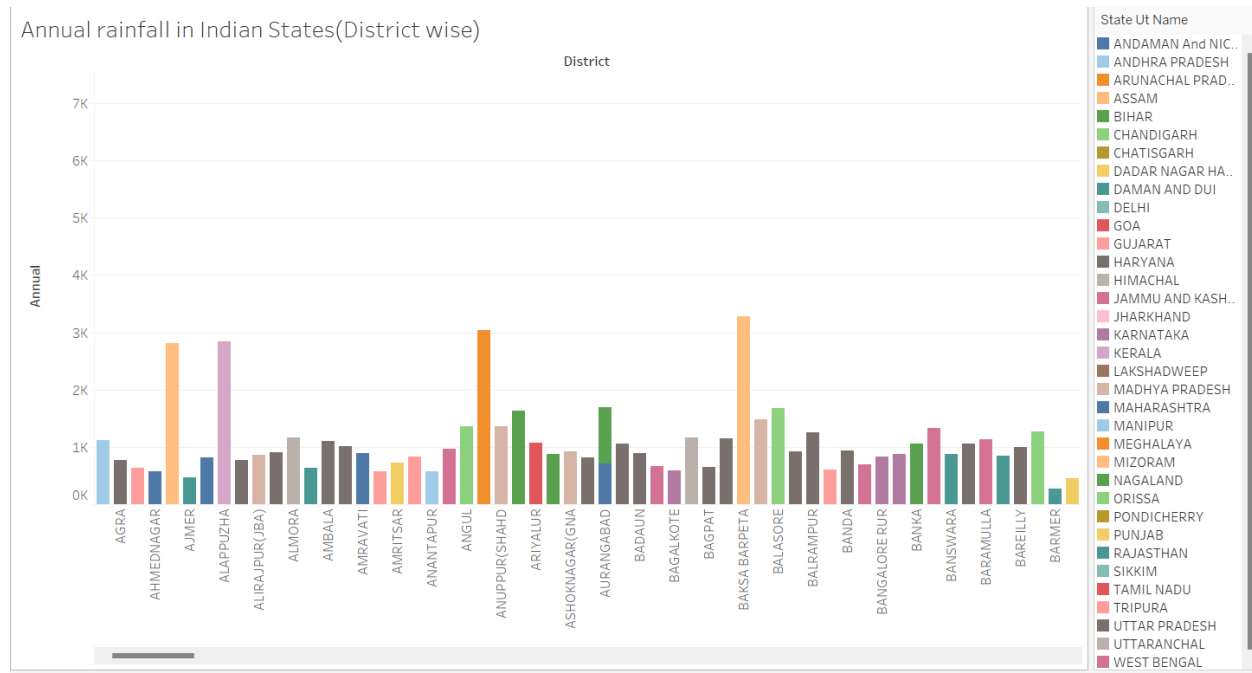
ARUNACHAL PRADESH

Data Visualization Sheets

1. Visualization of the annual rainfall received by each state using geospatial analysis. We can observe the region in the lighter shade of blue means those areas receive lesser rainfall in comparison to the darker coloured states ,which receive higher annual rainfall. The Legend shows the the color gradient for the gradual increase in rainfall received per annum in mm.

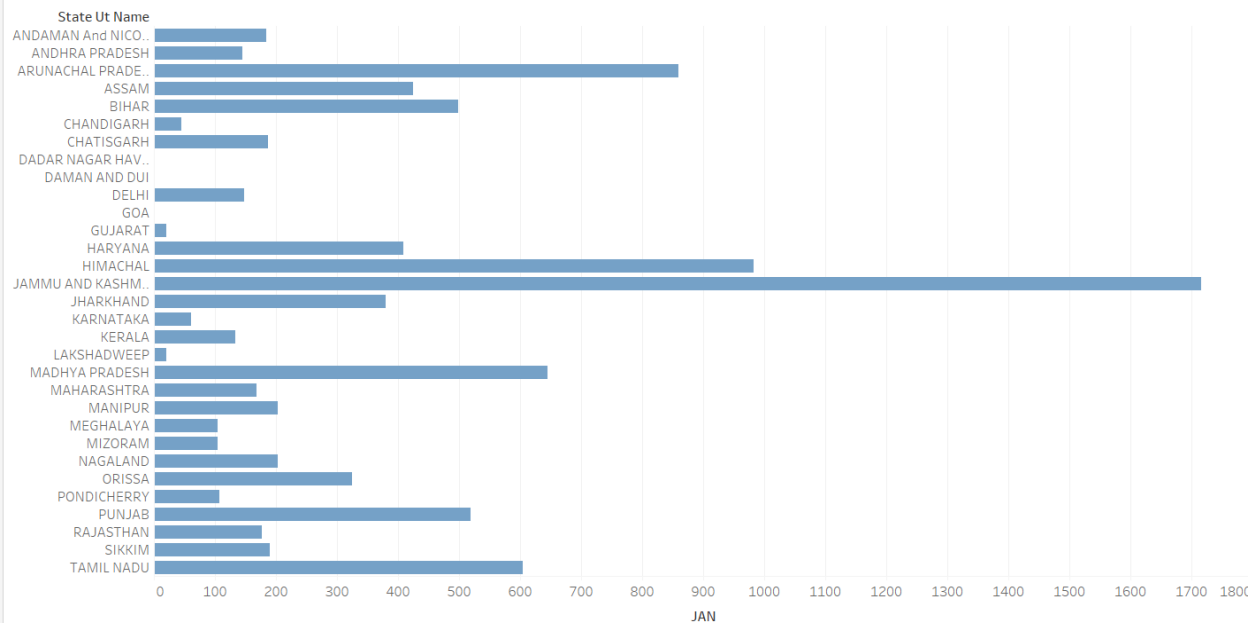


2.This visualization shows us the annual rainfall received by each district in the dataset annually measured in mm.The colouring of the bar graphs is done as per the states the districts belong to.This allows to give us a clear visual representation of how much each district of a state get and makes it easier to compare it with other district.When we select the filter by color on the right it will only show the district of the selected state,making further district wise comparison within state easier.



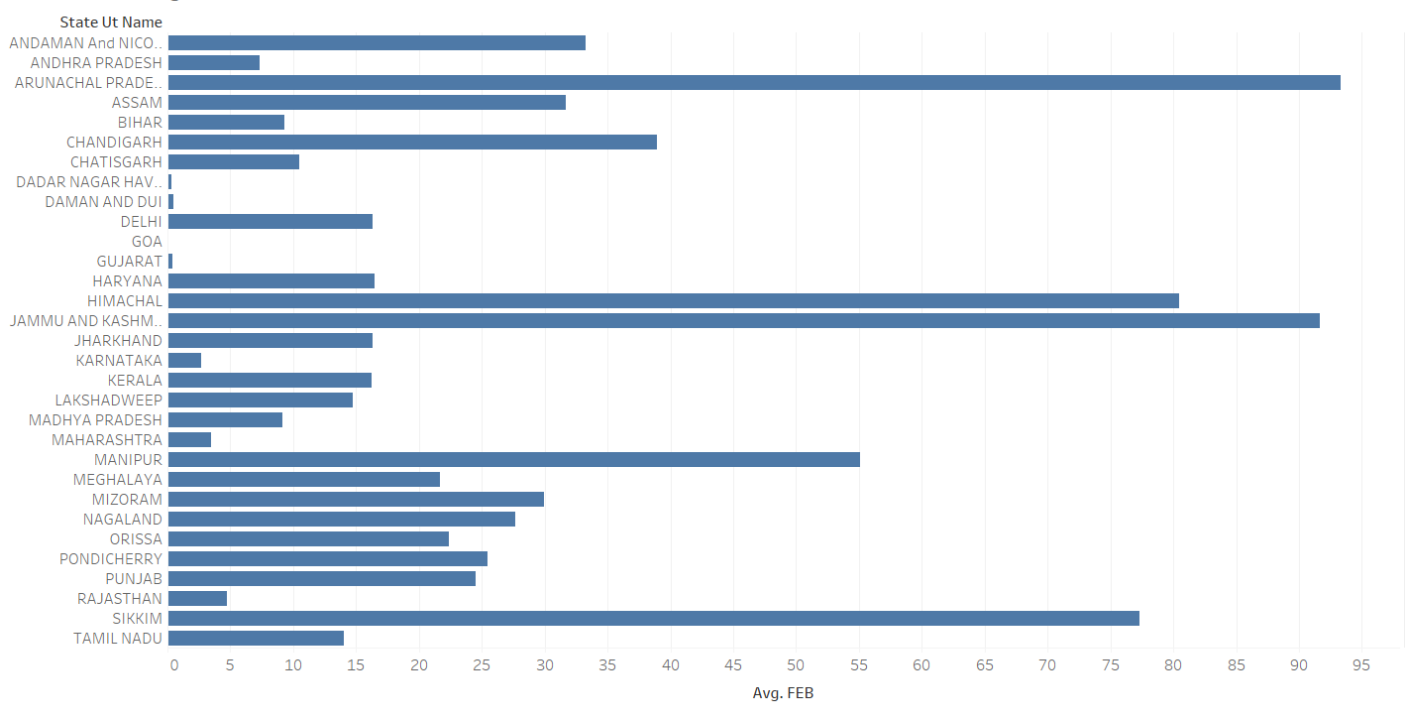
3.Analysis of rainfall received by states in the month of January using a horizontal bar plot. As we can observe from the plot on average the months that receive more rainfall on average in the month of January over the years 1901-2015.

Rainfall analysis across states for the month of January

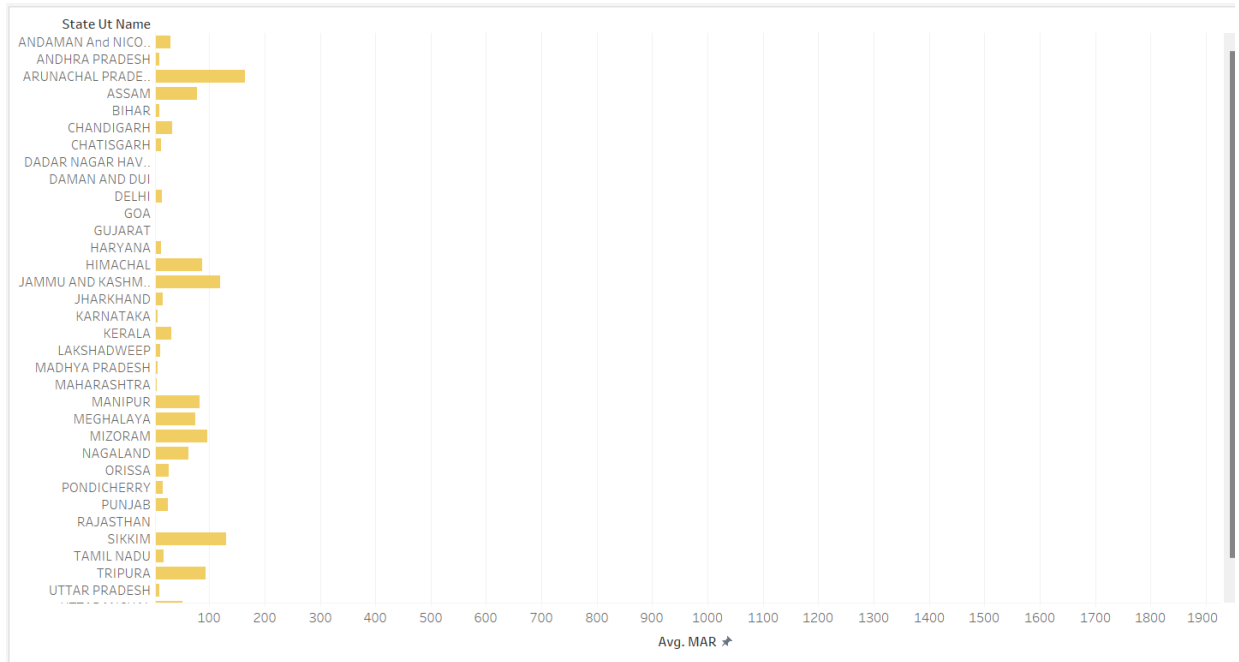


4. Analysis of rainfall received by states in the month of February using a horizontal bar plot. As we can observe from the plot on average the months that receive more rainfall on average in the month of February over the years 1901-2015.

Rainfall during the month of Feb

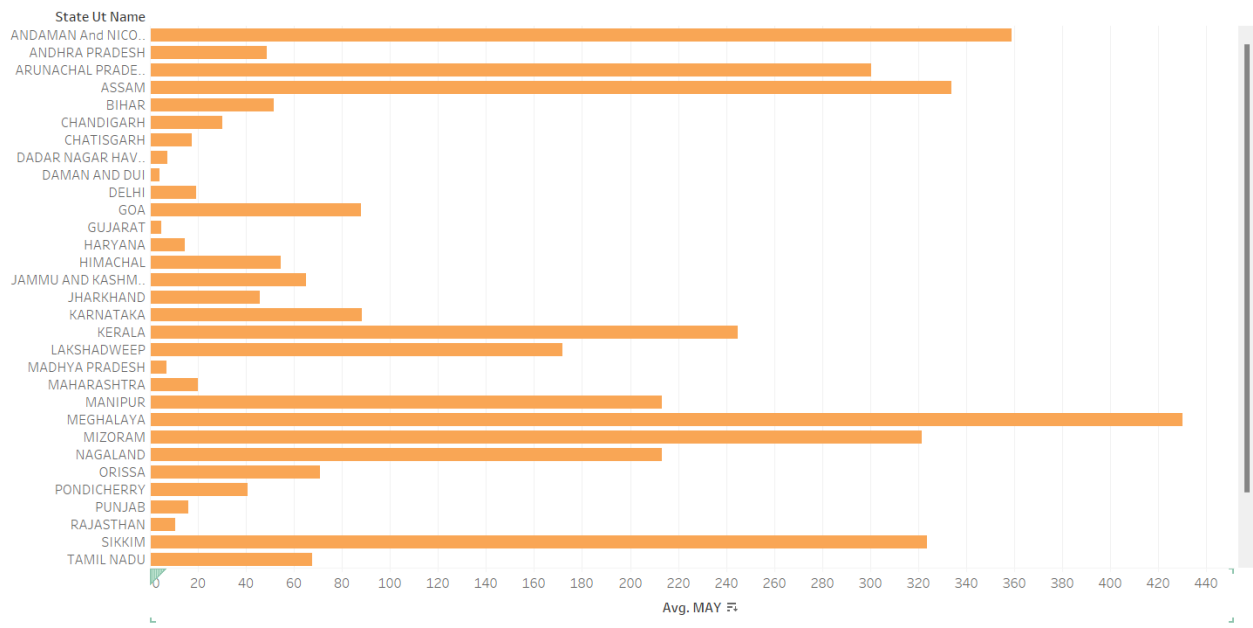


5. Analysis of rainfall received by states in the month of March using a horizontal bar plot. As we can observe from the plot on average the months that receive more rainfall on average in the month of March over the years 1901-2015.



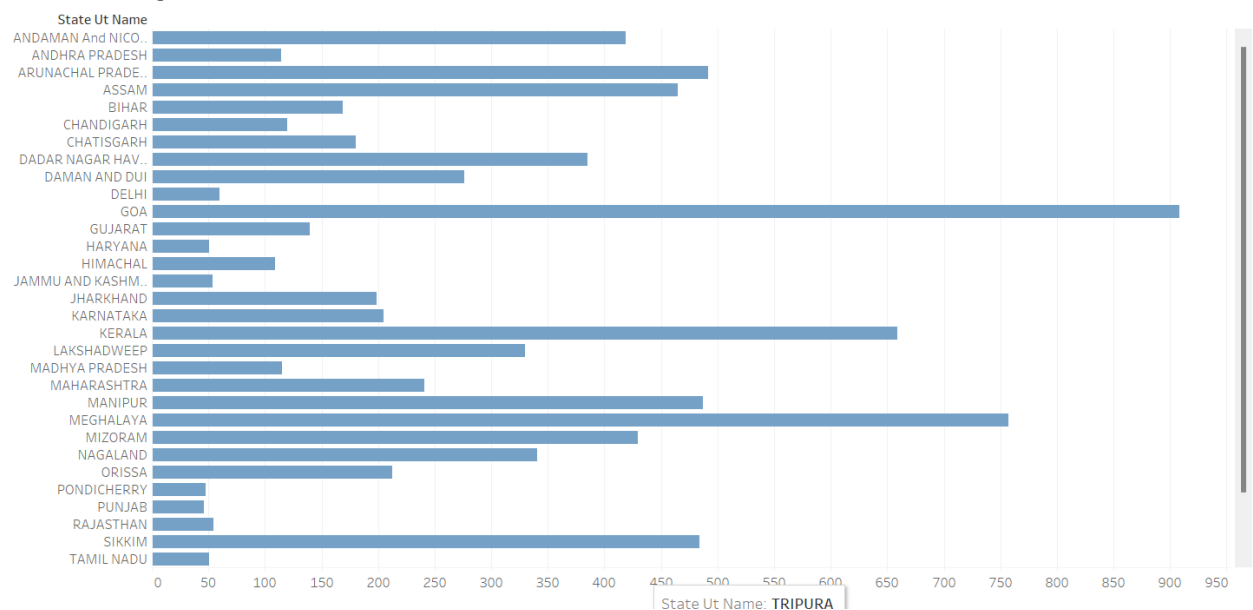
6. Analysis of rainfall received by states in the month of May using a horizontal bar plot. As we can observe from the plot on average the months that receive more rainfall on average in the month of May over the years 1901-2015.

Rainfall During May

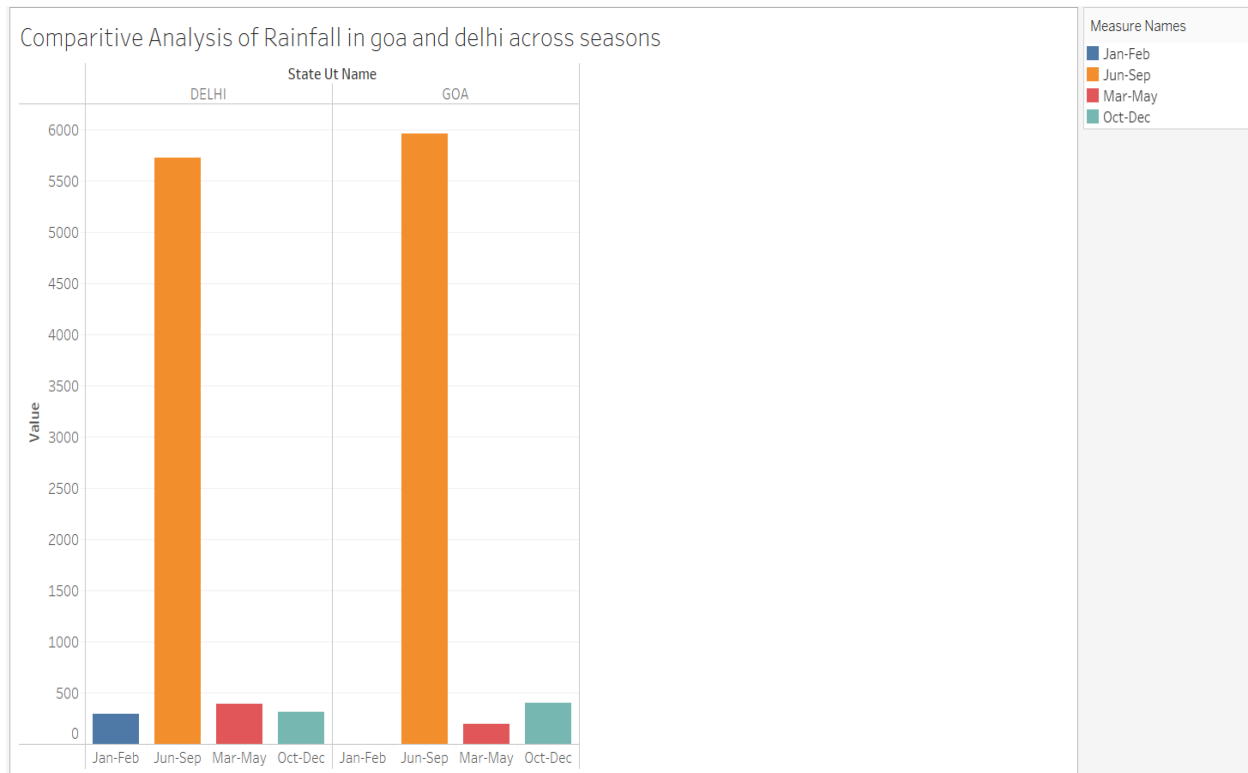


7. Analysis of rainfall received by states in the month of June using a horizontal bar plot. As we can observe from the plot on average the months that receive more rainfall on average in the month of June over the years 1901-2015.

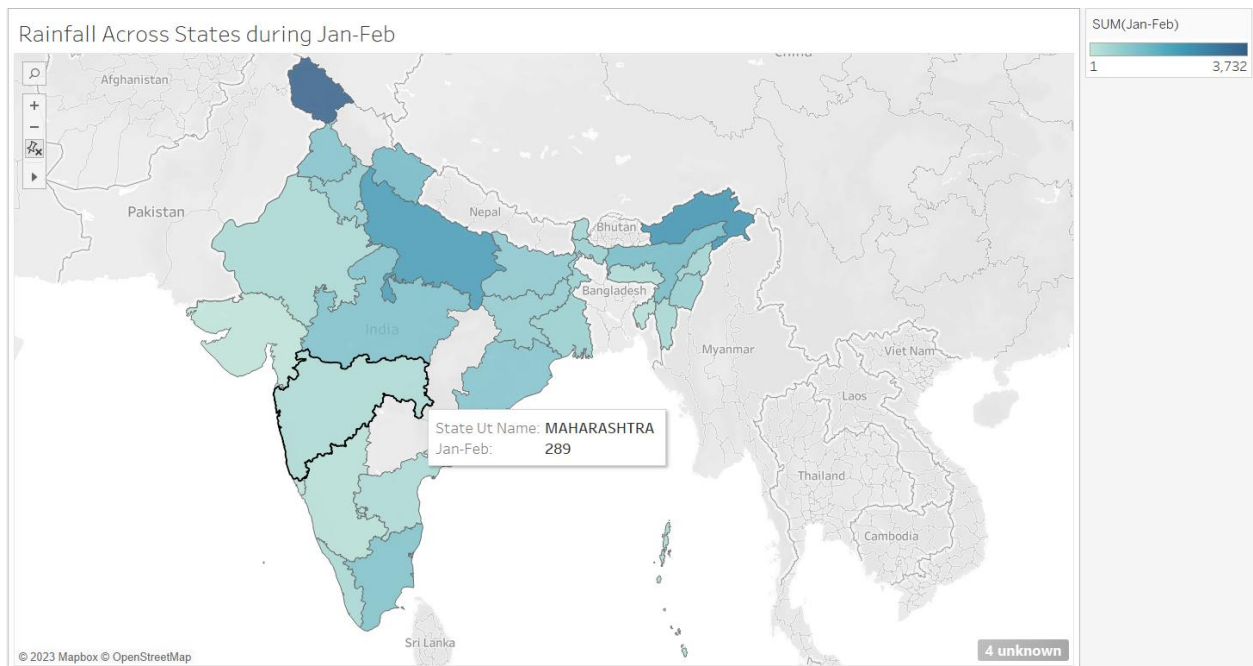
Rainfall During June



8. Here we have 2 sections, each belonging to Goa and Delhi respectively. Within these sections there are 3 colour-coded bar graphs that display the total rainfall received by the state during 1901-2015. We can see that the period between June to September receives maximum rainfall between 1901-2015. And that Delhi receives more rainfall in comparison to Goa during Jan-Feb.

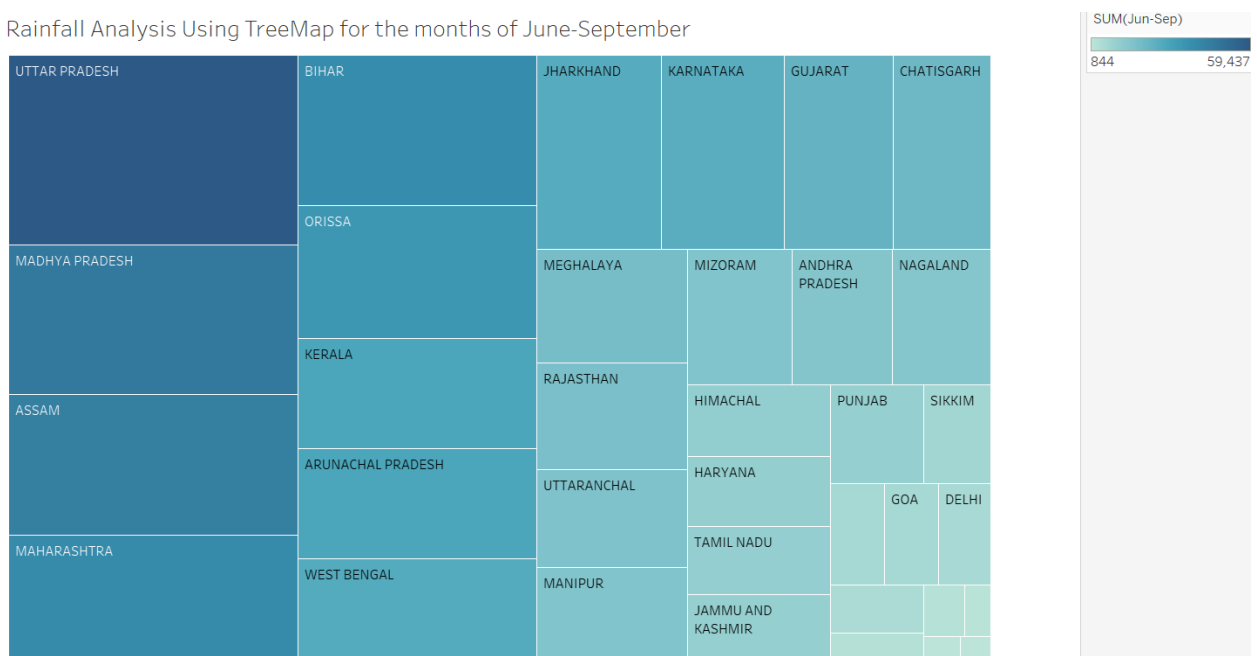


9. Analysis of rainfall received by states in the month of January using a Geospatial plot. As we can observe from the plot the states with higher rainfall during the month are darker shades of blue in comparison to lighter shaded states with less rainfall. We can see that northern region receives more rainfall during these months.

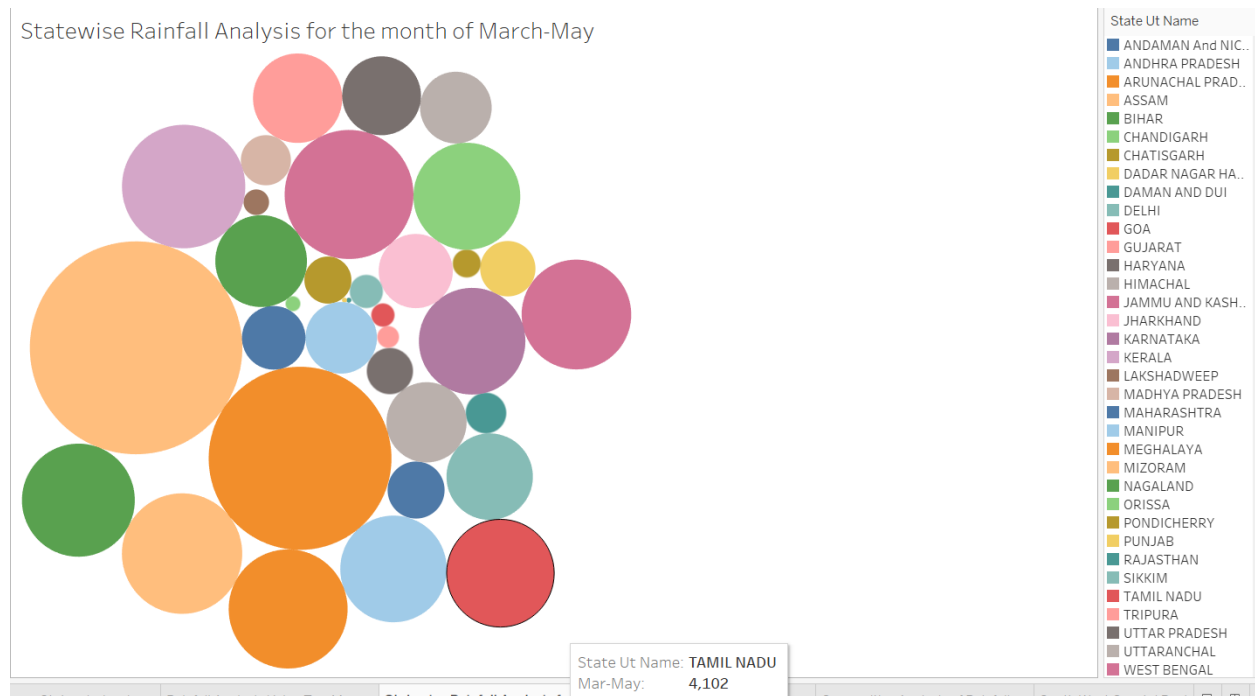


10.The visualized TreeMap gives us a hierarchy of states based on the rainfall received during the months of June-September. The states with higher rainfall are given bigger and darker coloured tiles in comparison to the states with less rainfall.

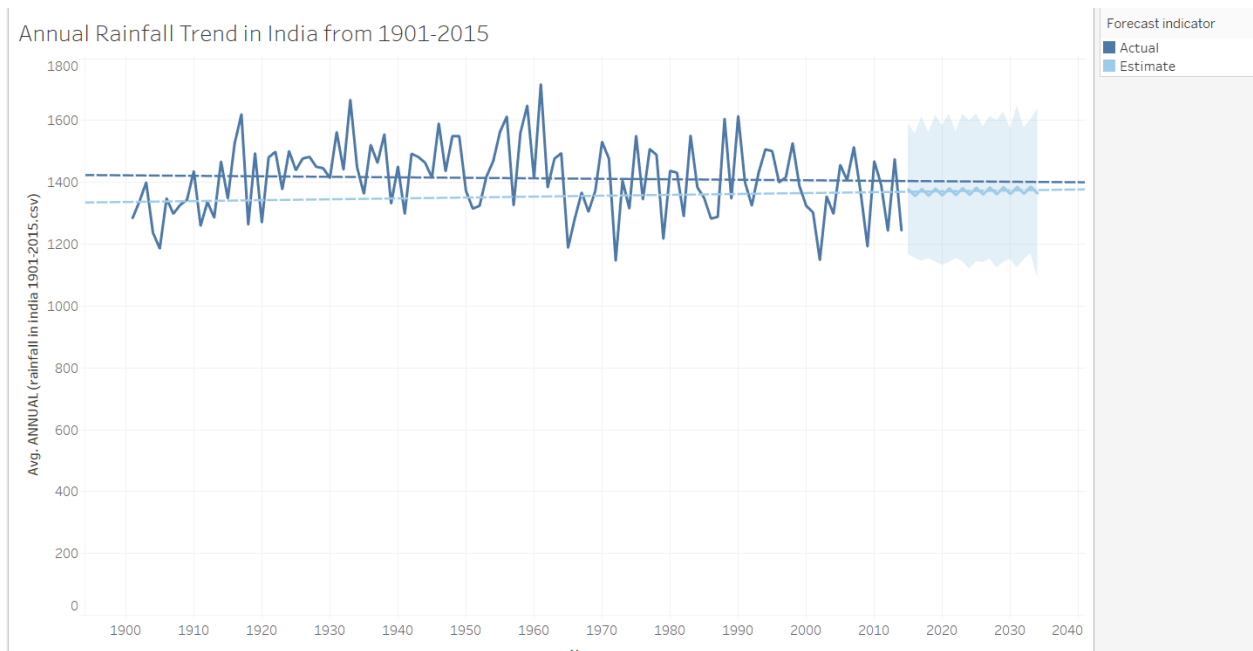
Rainfall Analysis Using TreeMap for the months of June-September



11. The visualized Bubble Char gives us a hierarchy of states based on the rainfall received during the months of March-May. The states with higher rainfall are given bigger bubbles in comparison to the states with less rainfall. Colors are awarded based on states. Highlight option is enabled to be selective with display.

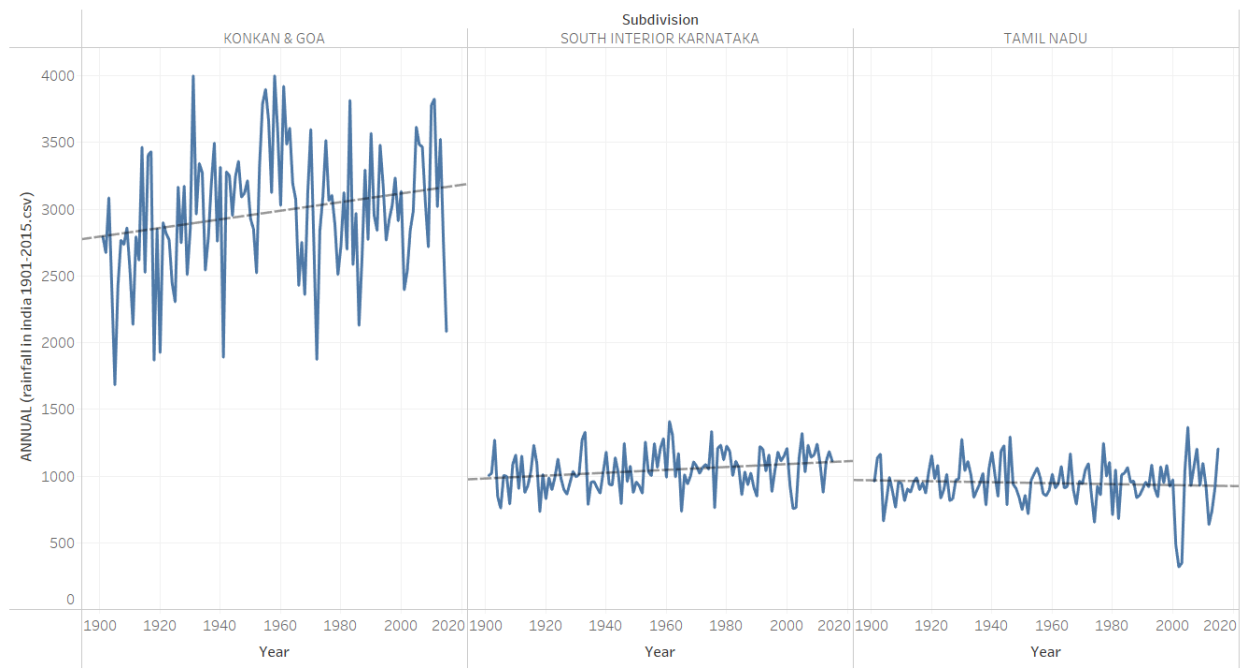


12. The visualization shows us the trend that the annual rainfall measure follows from 1901-2015 with the help of line chart. The peaks represent hikes and troughs depict lows. i.e. years where rainfall increased or drop. Forecasting has also been used to forecast predictive annual rainfall for the new 15 years 2021-2035.

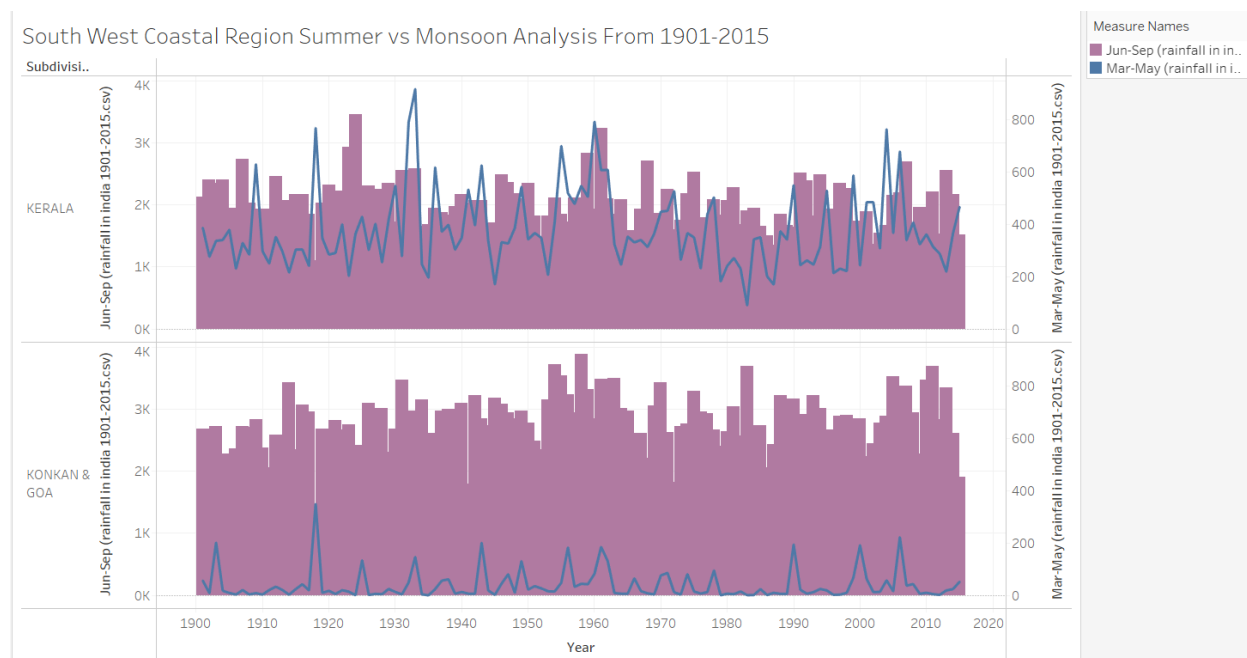


13.The following visualization allows to do comparative analysis of 3 states and the annual rainfall patterns in these states on the years from 1901-2015. I have made use of line charts and trend lines. This helps to generate a quick analysis of how the rainfall trend has been over the years. We can observe that the Annual rainfall in Konkan & Goa have an increase in the annual rainfall received from 1901 to 2021. The region of South Interior Karnataka has a very small gradual increase in the annual rainfall received. Whereas in the State of Tamil Nadu the rainfall trend has a gradual fall, signalling the decline in the total annual rainfall received over the years.

Comparative Analysis of Rainfall across 3 regions

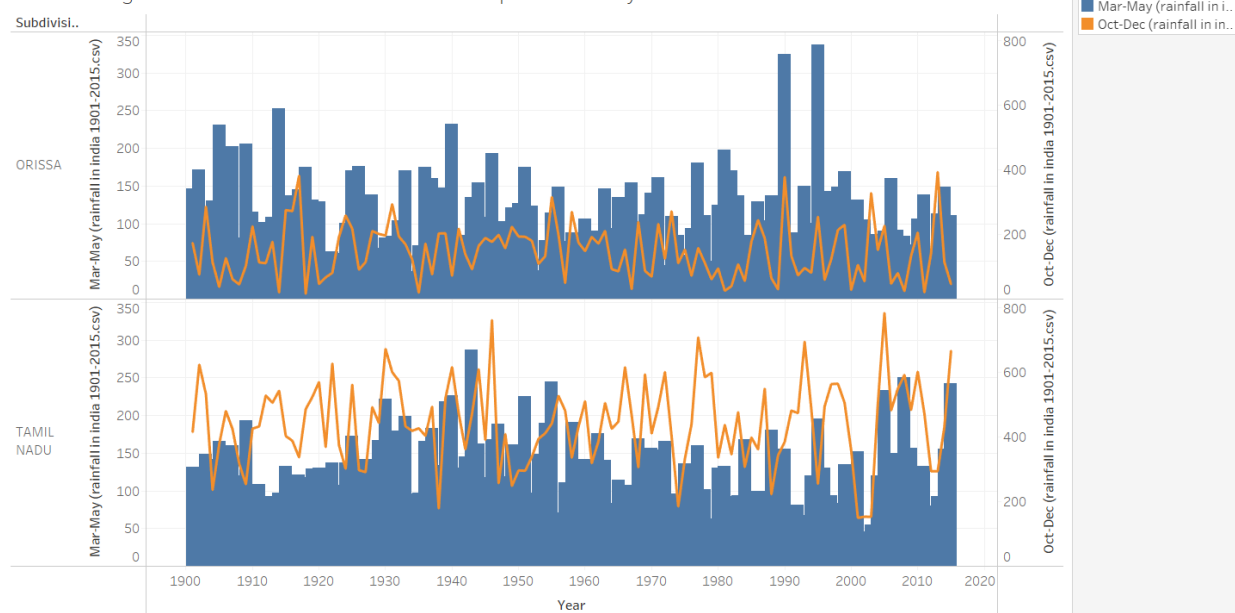


14. The following visualization allows to do comparative analysis of 2 states in the South West Region of India and the Monsoon rainfall patterns in these states over the years from 1901-2015. I have made use of line charts and bar graphs overlap. This helps to generate a quick analysis of how the rainfall trend has been during the Monsoon Season vs Non Monsoon Season. We can observe that the Annual rainfall in Kerala during Mar-May and Jun-Sept are almost similar but slightly higher during Jun-Sept, whereas in Goa the Months of June-Sept receive maximum rainfall and the month of Jan-Feb receive less.



15. The following visualization allows to do comparative analysis of 2 states in the Eastern Region of India and the Monsoon rainfall patterns in these states over the years from 1901-2015. I have made use of line charts and bar graphs overlap. This helps to generate a quick analysis of how the rainfall trend has been during the Monsoon Season vs Non Monsoon Season. We can observe that the Annual rainfall in Orissa during Mar-May is higher than during Oct-Dec, whereas in Tamil Nadu the Months of Oct-Dec receive maximum rainfall and the month of Mar-May receive less.

Eastern Region of India Summer Vs Monsoon Comparison Analysis from 1901-2015

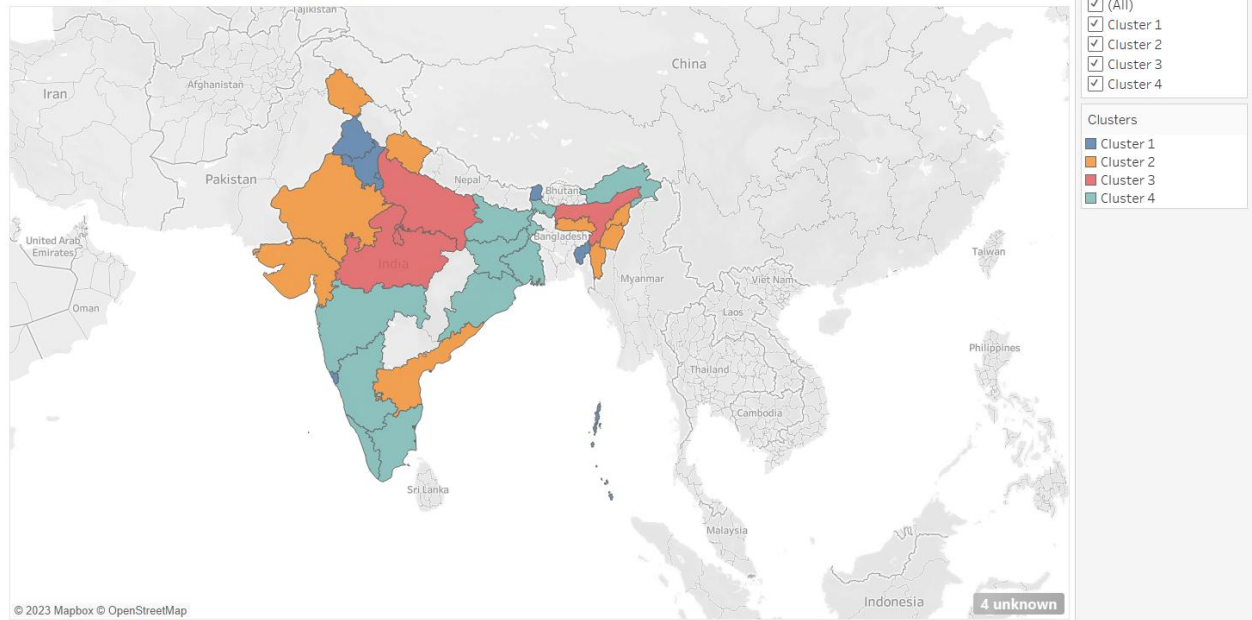


16. Clustering based on the annual rainfall received by the respective states. Clustering performed by assigning states with values that lie at minimum distance from centroid. The number of clusters defined in 4. the respective centroid measure for the clustering purpose are given in the screenshot below:

Number of Clusters: 4
 Number of Points: 35
 Between-group Sum of Squares: 2.2407
 Within-group Sum of Squares: 0.16927
 Total Sum of Squares: 2.4099

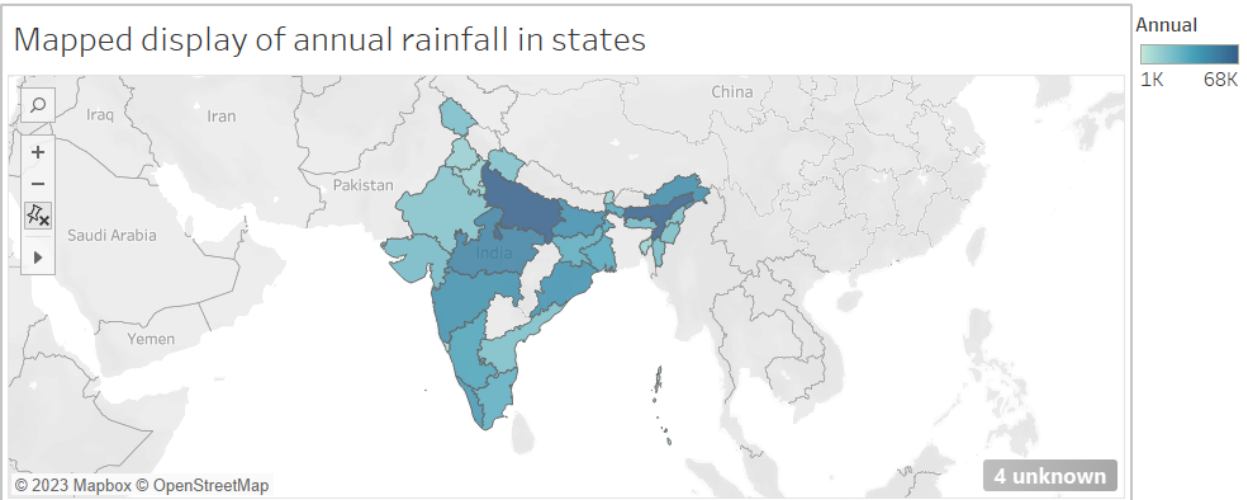
Clusters	Number of Items	Centers Sum of Annual
Cluster 1	12	6899.3
Cluster 2	11	21849.0
Cluster 3	3	61907.0
Cluster 4	9	39395.0
Not Clustered	0	

Clustering based on annual rainfall

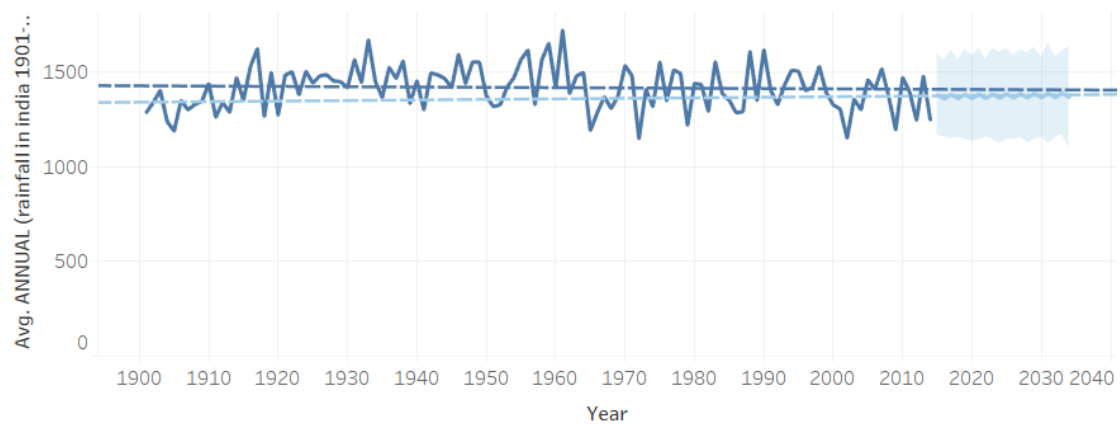


Data Visualization Dashboards

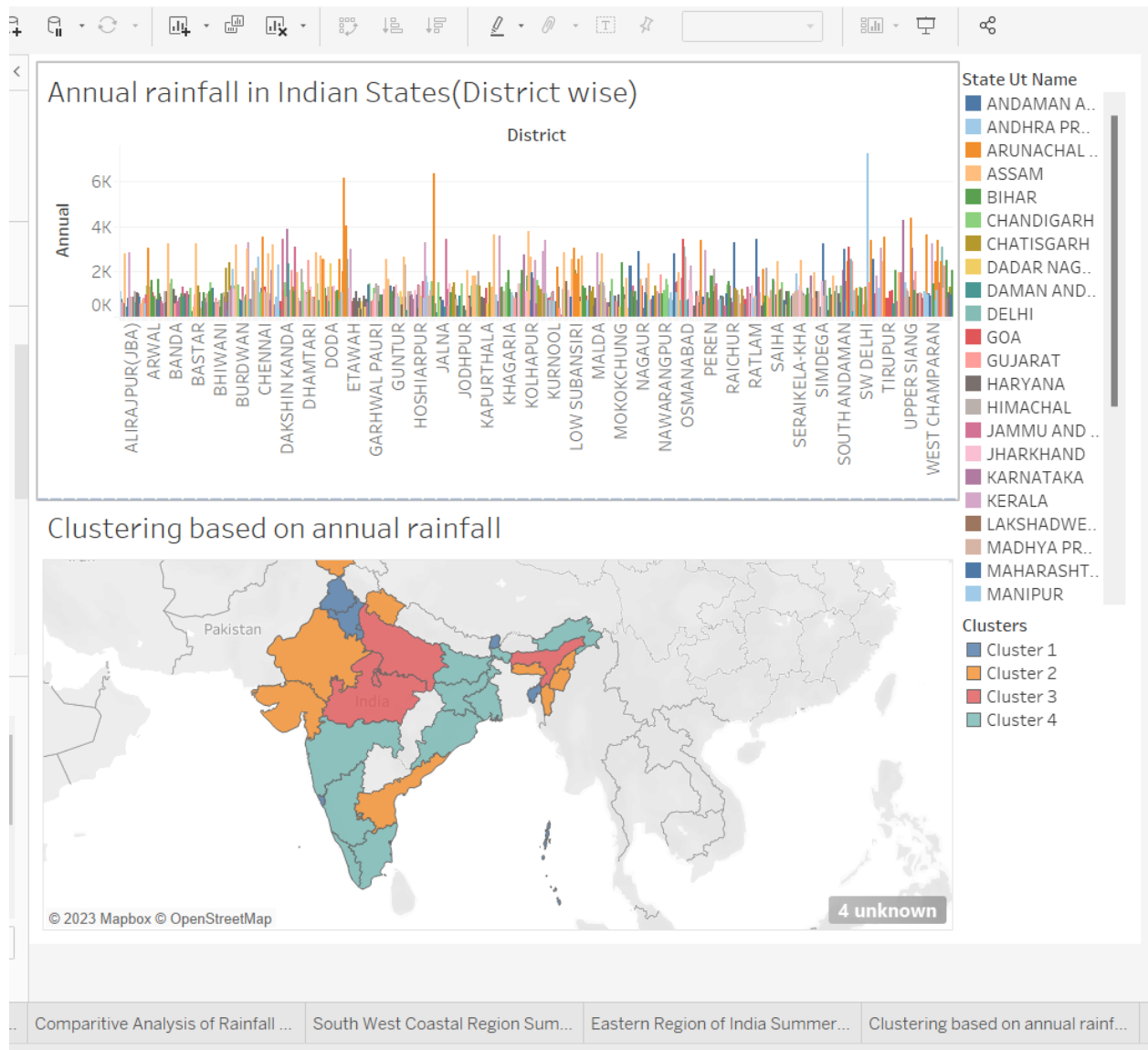
1.



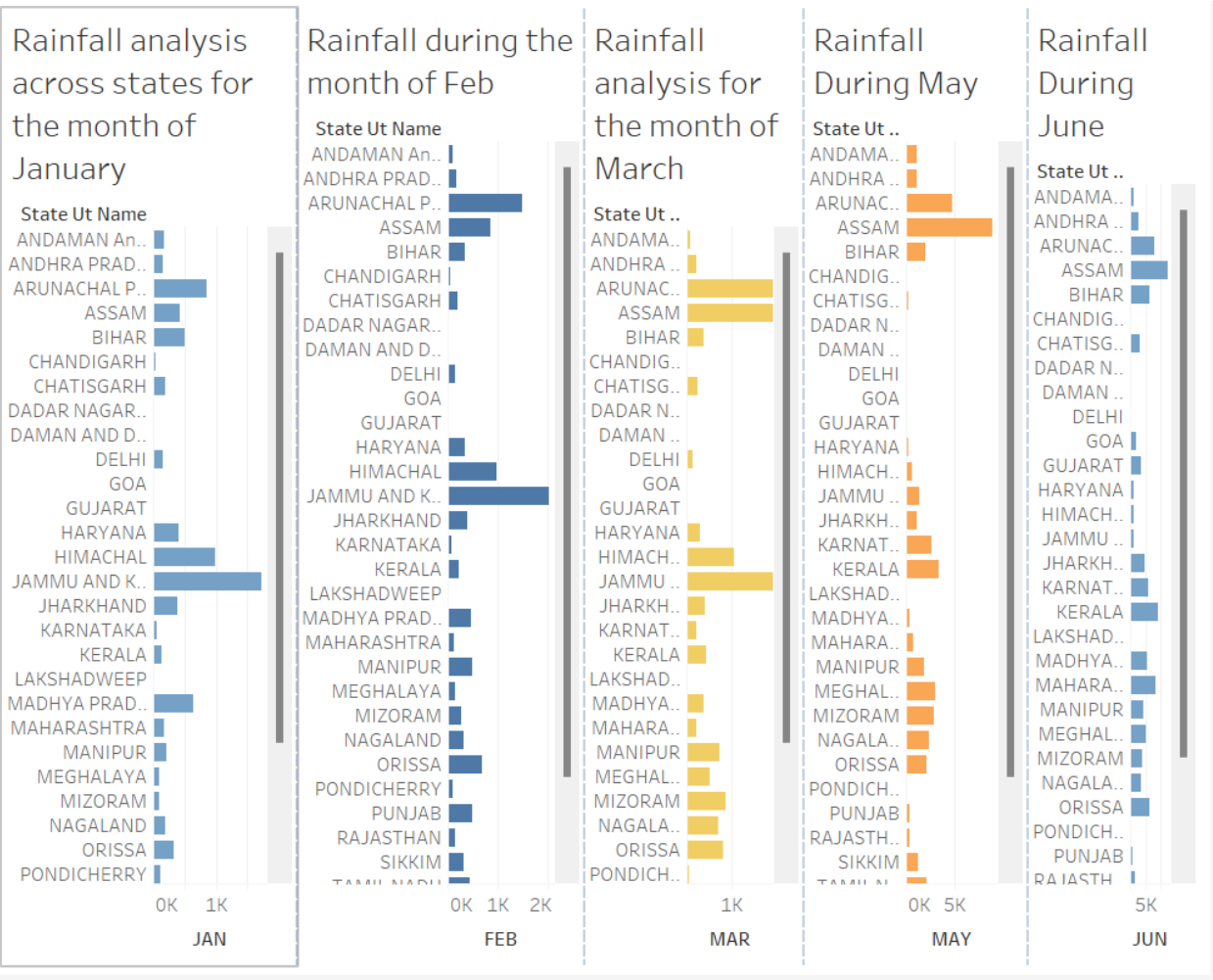
Annual Rainfall Trend in India from 1901-2015



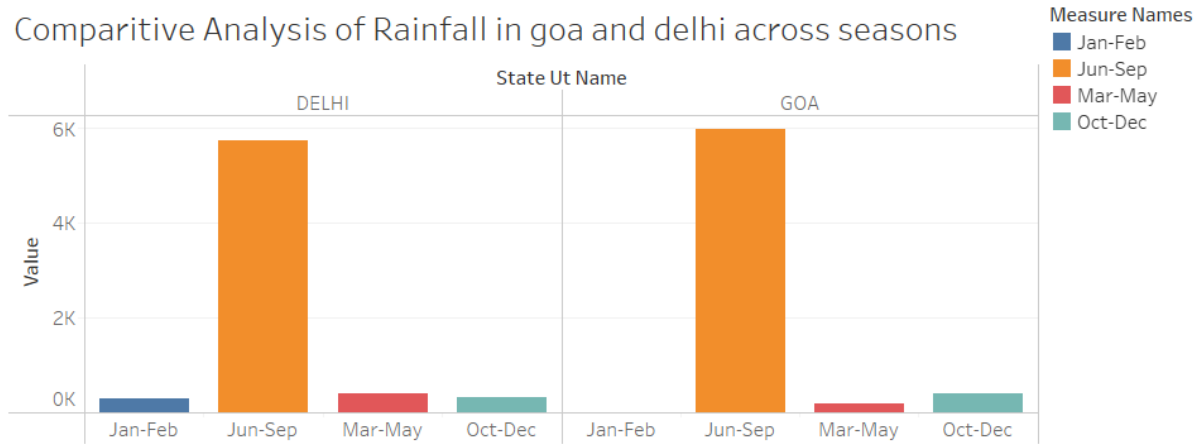
2.



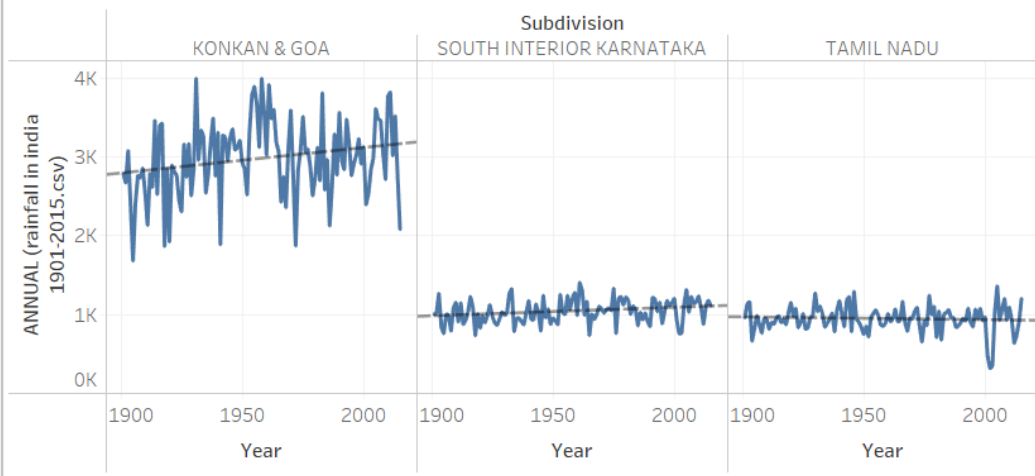
3.



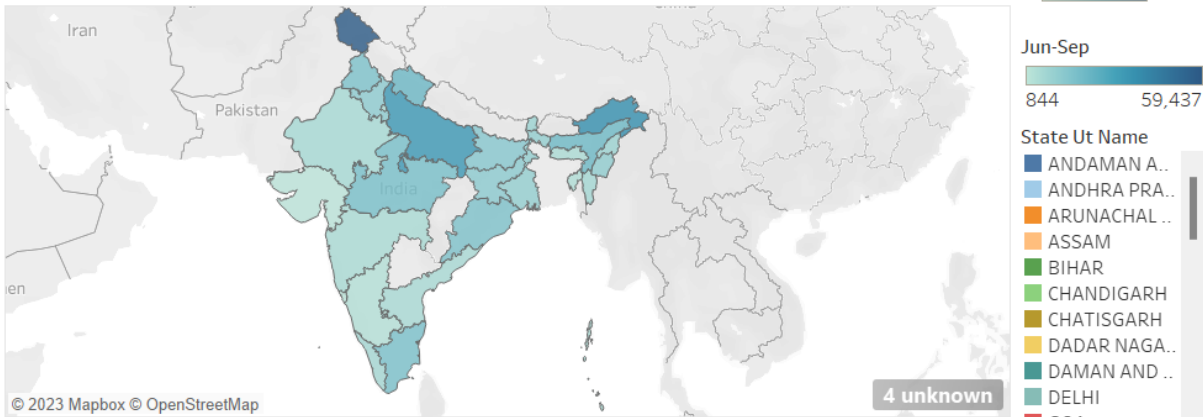
Comparitive Analysis of Rainfall in goa and delhi across seasons



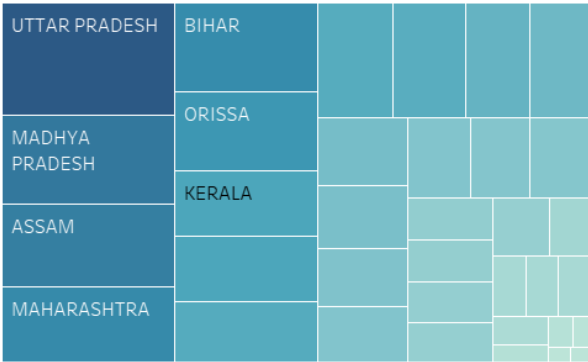
Comparitive Analysis of Rainfall across 3 regions



Rainfall Across States during Jan-Feb



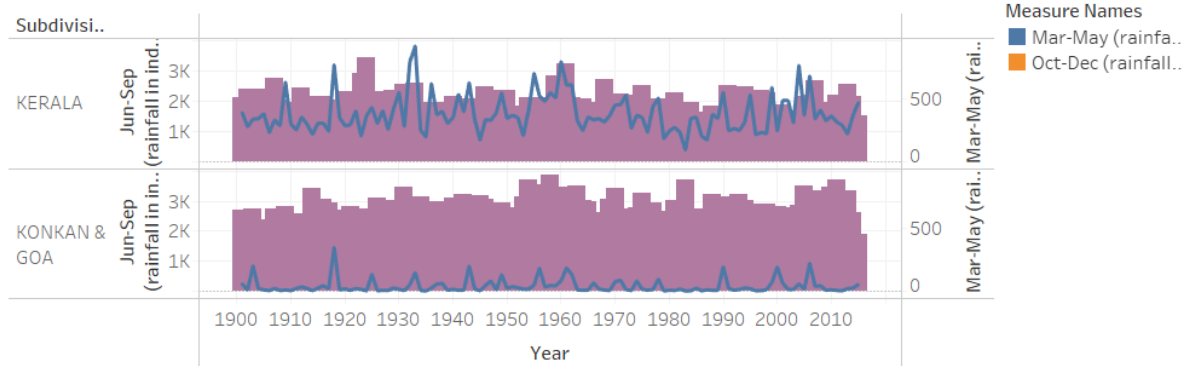
Rainfall Analysis Using TreeMap for the months of June-September



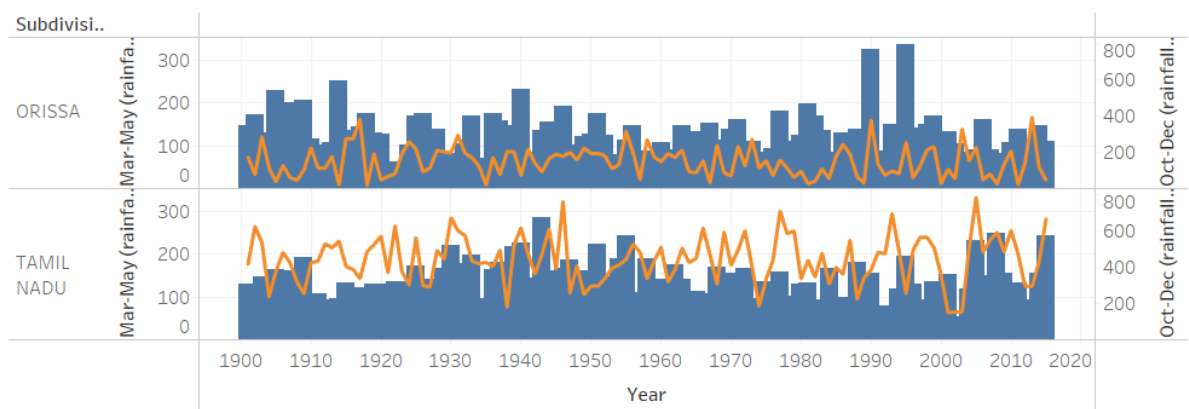
Statewise Rainfall Analysis for the month of March-May



South West Coastal Region Summer vs Monsoon Analysis From 1901-2015



Eastern Region of India Summer Vs Monsoon Comparison Analysis from 1901-2015



Clustering based on annual rainf... ☐ Overall Statistics of Rainfall in I... ☐ District and state wise ananly... ☐ Observing Gradual decline du... ☐