

About The Project

This project is about the concepts which we have learnt in R programming. This project mainly focus on ggplot() function and analyzing and visualizing the data on Forest Cover of the states of India and it's relation with the amount of annual rainfall it receives. This project uses various visualization tools such as scatter plot, regression lines etc. to draw inferences.

In this project we have taken a sample .csv file name as rainfall.csv and imported it into RStudio, then we have applied predefined functions and library and drew the charts.

The data has been collected from various websites like The Forest Survey of India and mydata.gov.in.

Plotting the graphs from rainfall.csv

- Setting of Working Directory

```
setwd ("D:/rp")
```

- Reading of .csv file

```
data <- read.csv("rainfall.csv")
```

- Installing ggplot package

```
install.packages("ggplot2")
```

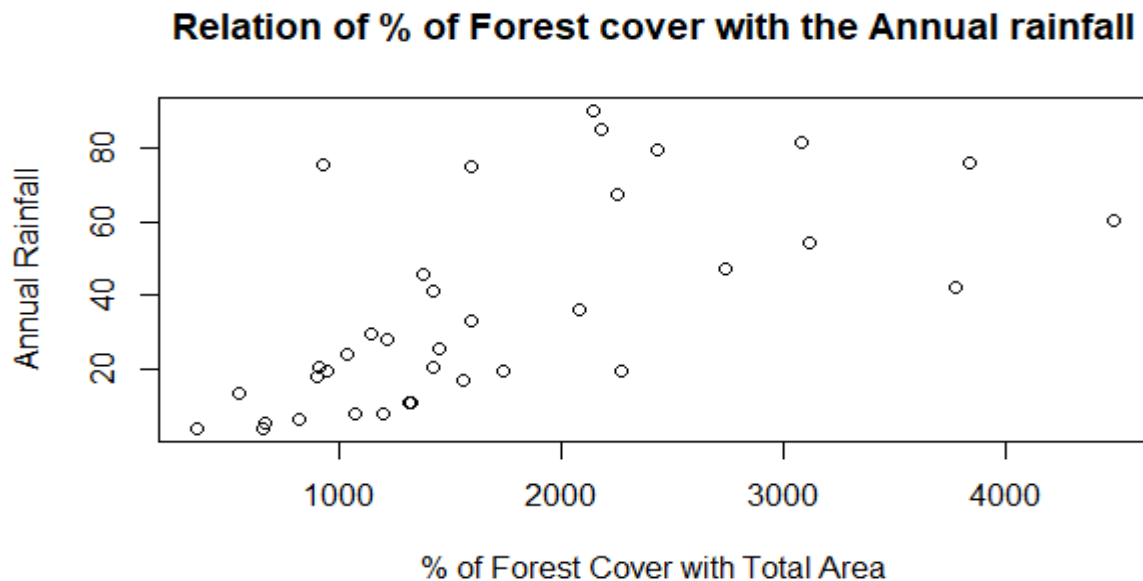
this package is important for plotting graphs and charts few of them will be shown below.

- Using ggplot() library

```
library(ggplot2)
```

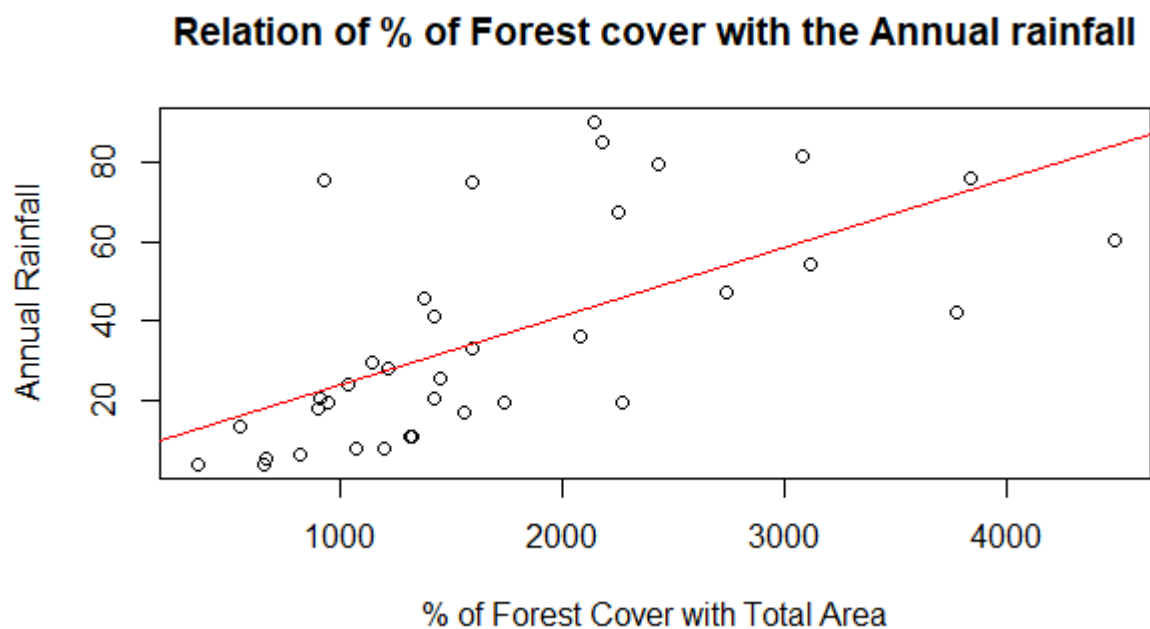
1. Scatter plot

```
plot(data$annual.rainfall,data$percent,main="Relation of % of  
Forest cover with the Annual rainfall",xlab="% of Forest Cover with  
Total Area",ylab="Annual Rainfall")
```



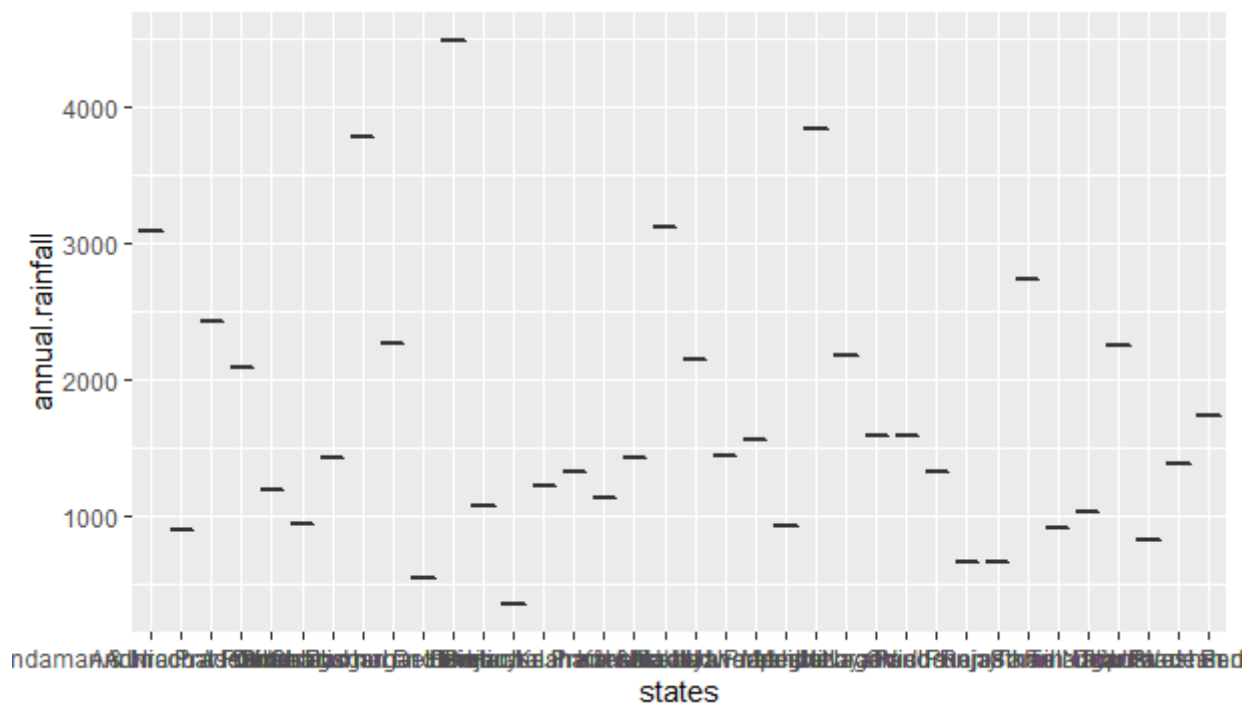
2. Now, adding the line of best fit

```
abline(lm(percent~annual.rainfall,data=data),col="red")
```



3. Boxplot

```
ggplot(data, aes(x=States, y=percent))+ geom_boxplot()
```



4. Some Quantitative Data

- **Minimum**

```
min(data$annual.rainfall)
```

- **Maximum**

```
max(data$annual.rainfall)
```

- **Mean**

```
mean(data$annual.rainfall)
```

```
mean(data$percent)
```

- **Median**

```
median(data$annual.rainfall)
```

```
median(data$percent)
```

- **Quantile**

`quantile(data$annual.rainfall,0.25)`

`quantile(data$annual.rainfall, 0.75)`

- **Standard Deviation And Variance**

`sd(data$annual.rainfall)`

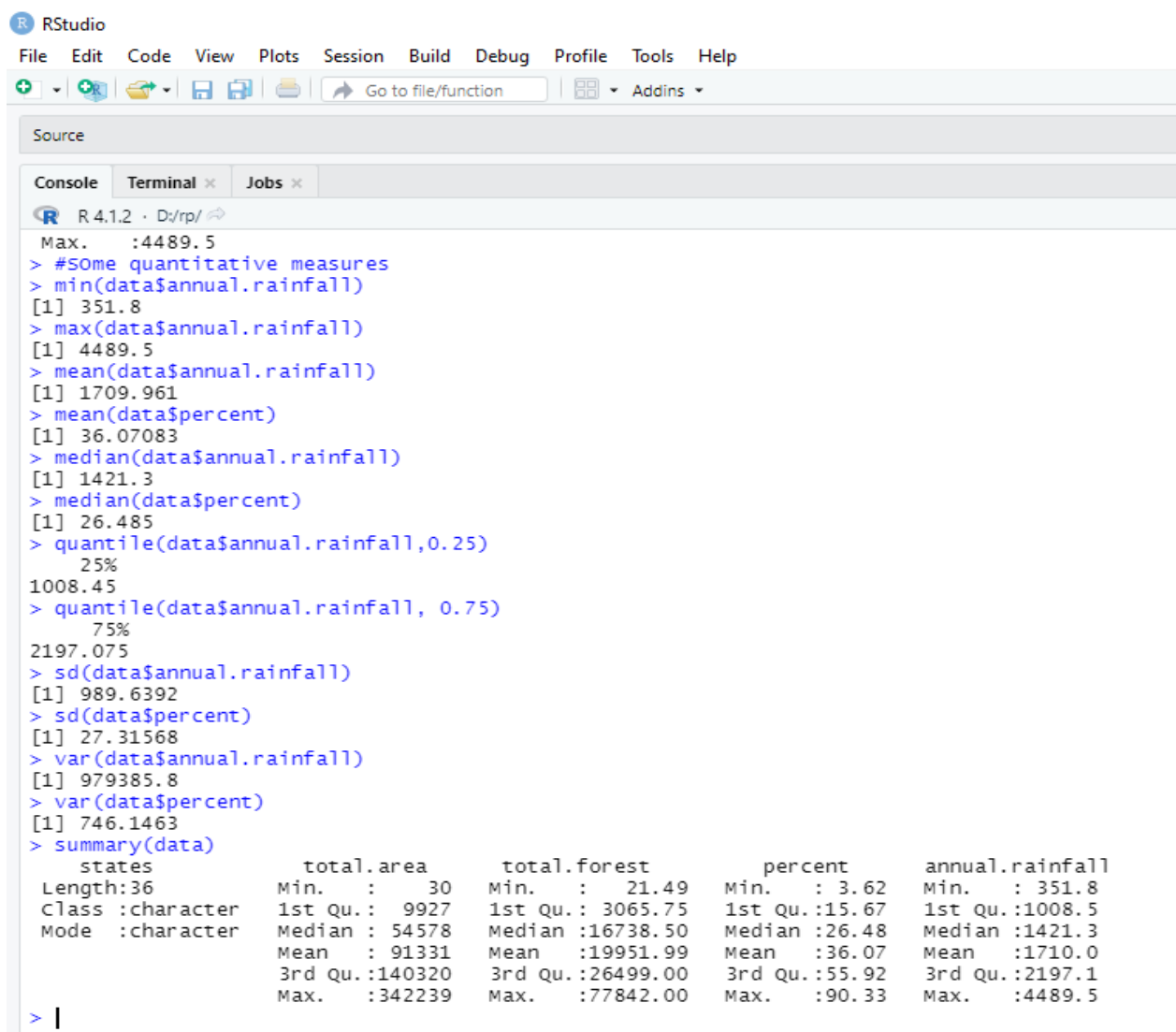
`sd(data$percent)`

`var(data$TotalProfit)`

`var(data$percent)`

- **Summary**

`summary(data)`

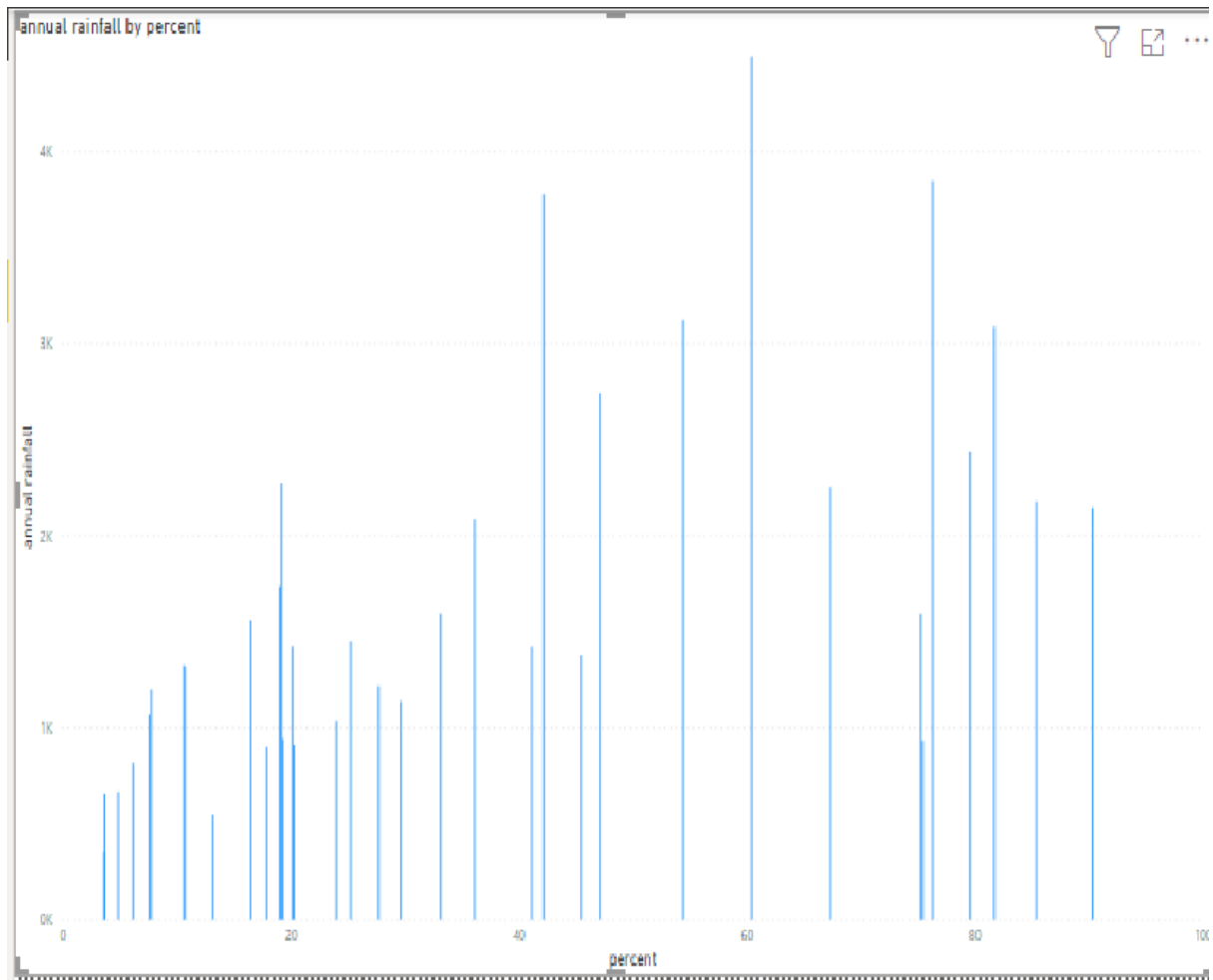


The screenshot shows the RStudio interface with the console pane active. The console displays the results of several R commands. The first command, `Max. :4489.5`, shows the maximum value of the `annual.rainfall` variable. Subsequent commands calculate the minimum, maximum, mean, and median for both `annual.rainfall` and `percent`. The `quantile` function is used to find the 25th and 75th percentiles for `annual.rainfall`. The `sd` (standard deviation) and `var` (variance) functions are applied to both variables. Finally, the `summary(data)` command is executed, resulting in a summary table for the `states` variable.

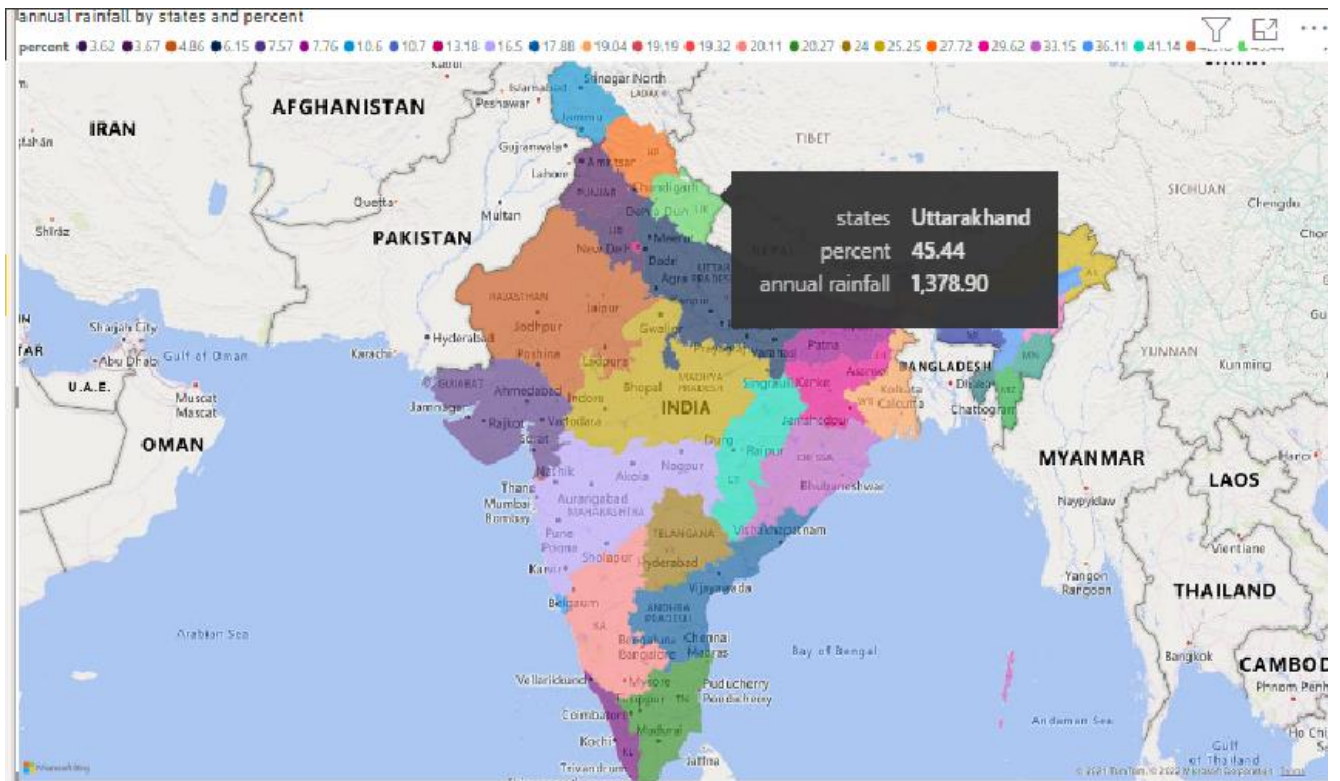
```
R 4.1.2 · D:/rp/ ↗
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal x Jobs x
R 4.1.2 · D:/rp/ ↗
Max. :4489.5
> #Some quantitative measures
> min(data$annual.rainfall)
[1] 351.8
> max(data$annual.rainfall)
[1] 4489.5
> mean(data$annual.rainfall)
[1] 1709.961
> mean(data$percent)
[1] 36.07083
> median(data$annual.rainfall)
[1] 1421.3
> median(data$percent)
[1] 26.485
> quantile(data$annual.rainfall,0.25)
25%
1008.45
> quantile(data$annual.rainfall, 0.75)
75%
2197.075
> sd(data$annual.rainfall)
[1] 989.6392
> sd(data$percent)
[1] 27.31568
> var(data$annual.rainfall)
[1] 979385.8
> var(data$percent)
[1] 746.1463
> summary(data)
states      total.area      total.forest      percent      annual.rainfall
Length:36   Min. :    30   Min. :  21.49   Min. :  3.62   Min. :  351.8
Class :character 1st Qu.:  9927 1st Qu.: 3065.75 1st Qu.:15.67 1st Qu.:1008.5
Mode :character Median: 54578 Median:16738.50 Median:26.48 Median:1421.3
Mean : 91331 Mean :19951.99 Mean :36.07 Mean :1710.0
3rd Qu.:140320 3rd Qu.:26499.00 3rd Qu.:55.92 3rd Qu.:2197.1
Max. :342239 Max. :77842.00 Max. :90.33 Max. :4489.5
> |
```

Some visualizations using Microsoft Power BI

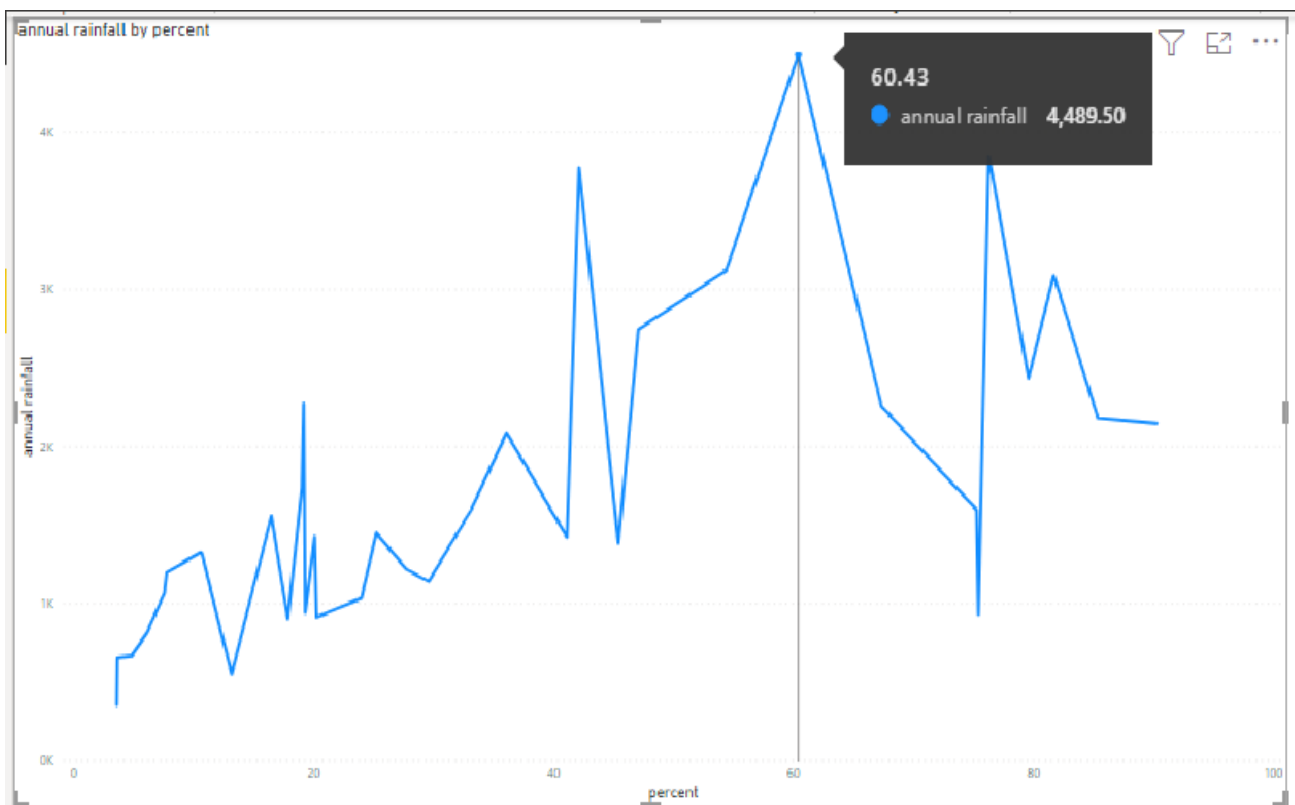
5. Stacked Column Chart



6. Filled Map



7. Line Chart



Inference drawn from the above statistical visualizations :

- 1.** States which have a higher percent of their total geographical area under forest cover receive higher annual rainfall.
- 2.** Coefficient of correlation between percent of area under forest cover and total annual rainfall comes out to be 0.629632 which shows a positive correlation, i.e. as the percent of forest area increases so does the total annual rainfall.
- 3.** Although, there are some outliers as can be seen in the line of best fit scatter plot, there are many geographical reasons like location, extent and other climatic factors.