# Exploratory Data Analysis on Water Reservoirs Data in India

Problem Statement:

- India is the Second largest population country in the world. Water is the essential need of every routine work for manufacturing Industries, agriculture and other domestic purposes. India facing the water scarcity in the summer season due to the shortage of water. So, the storage of water in rainy season and in winter season is the essential one to be carried.
- Frequent Monitoring of water storage reservoirs is the important one to be carried by the government to avoid the water scarcity in the summer season.
- This analysis is very useful for the Government to take other necessary measures to save the water for the future use.

Objective of Analysis:

- Finding the unnoticed loss of water from the Reservoirs.
- Identifying which category of Reservoirs has more loss of water

Assumptions:

- The dataset is accurate, relevant, no missing values and containing no outliers in it.
- The Storage capacity and water level is positively correlated to each other.
- It may not be correlated to different storage category of the reservoirs.

Description of the dataset:

The dataset is taken from https://data.gov.in/resource/daily-data-reservoir-level-central-water-commission-cwc-agency-during-february-2023/ . The dataset contains a data about the water reservoirs in India and their attributes are explained below.

- Reservoir name
- Basin
- Subbasin
- Agency name
- Date
- Year
- Month
- Full_reservoir_level
  Reservoirs full storage capacity in Meters

- Live_capacity_frl
  Live full storage water capacity of water in Billion Cubic Meter
- Storage
  Present water capacity in Billion Cubic Meter
- Level
  Present water level in meters

```
#Importing the packages for the analysis
library(dplyr)

##
## Attaching package: ‘dplyr’

## The following objects are masked from ‘package:stats’:
##
##     filter, lag

## The following objects are masked from ‘package:base’:
##
##     intersect, setdiff, setequal, union

library(lattice)

## Warning: package ‘lattice’ was built under R version 4.2.3

library(markdown)

#Loading the dataset
data=read.csv(“reservoir_waterlevel.csv”)

#Internal structure of the dataset
str(data)

## ‘data.frame’:    2856 obs. Of  11 variables:
##  $ Reservoir_name      : chr  “Aliyar Reservoir” “Aliyar Reservoir”
“Aliyar Reservoir” “Aliyar Reservoir” …
##  $ Basin               : chr  “West flowing rivers from Tadri to
Kanyakumari Basin” “West flowing rivers from Tadri to Kanyakumari Basin”
“West flowing rivers from Tadri to Kanyakumari Basin” “West flowing rivers
from Tadri to Kanyakumari Basin” …
##  $ Subbasin            : chr  “Varrar and others” “Varrar and others”
“Varrar and others” “Varrar and others” …
##  $ Agency_name         : chr  “CWC” “CWC” “CWC” “CWC” …
##  $ Date                : chr  “2023-02-01” “2023-02-02” “2023-02-03”
“2023-02-04” …
##  $ Year                : int  2023 2023 2023 2023 2023 2023 2023 2023 2023
2023 …
```

```
##  $ Month               : int  2 2 2 2 2 2 2 2 2 …
##  $ Full_reservoir_level: num  320 320 320 320 320 …
##  $ Live_capacity_FRL   : num  0.095 0.095 0.095 0.095 0.095 0.095 0.095
0.095 0.095 0.095 …
##  $ Storage             : num  0.0579 0.0568 0.0559 0.0549 0.0537 …
##  $ Level               : num  314 314 314 313 313 …

#Statistical summary of the dataset
summary(data)

##  Reservoir_name        Basin             Subbasin         Agency_name
##  Length:2856        Length:2856        Length:2856        Length:2856
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##      Date                Year          Month   Full_reservoir_level
##  Length:2856        Min.   :2023   Min.   :2   Min.   :  55.0
##  Class :character   1st Qu.:2023   1st Qu.:2   1st Qu.: 169.0
##  Mode  :character   Median :2023   Median :2   Median : 359.7
##                     Mean   :2023   Mean   :2   Mean   : 398.5
##                     3rd Qu.:2023   3rd Qu.:2   3rd Qu.: 594.4
##                     Max.   :2023   Max.   :2   Max.   :1002.8
##
##  Live_capacity_FRL    Storage           Level
##  Min.   :0.006     Min.   :0.0010   Min.   : 45.97
##  1st Qu.:0.288     1st Qu.:0.1760   1st Qu.:186.23
##  Median :0.664     Median :0.3674   Median :354.57
##  Mean   :1.409     Mean   :0.7399   Mean   :398.57
##  3rd Qu.:1.711     3rd Qu.:0.8692   3rd Qu.:550.35
##  Max.   :9.745     Max.   :6.3550   Max.   :972.19
##                    NA's   :816      NA's   :816

#Locating the NA values in the dataset
which(is.na(data$Storage))

##   [1]   24   48   51   52   53   54   55   57   58   59   60   61   62
64   65
##  [16]   66   67   68   69   71   72   73   75   76   77   78   79   80
82   83
##  [31]   84   85   86   87   89   90   91   92   93   94   96  120  144
168  190
##  [46]  192  216  217  219  220  221  222  223  224  226  227  228  229
230  231
##  [61]  233  234  235  236  237  238  240  264  265  267  268  269  270
271  272
##  [76]  274  275  276  277  278  279  281  282  283  284  285  286  288
312  314
##  [91]  315  316  317  318  319  321  322  323  324  325  326  327  329
```

```
330  331
## [106]  332  334  335  336  338  339  340  341  342  343  345  346  347
348  349
## [121]  350  351  353  354  355  356  358  359  360  384  386  394  395
396  397
## [136]  398  399  408  409  432  433  435  436  437  438  439  440  442
443  444
## [151]  445  446  447  449  450  451  452  453  454  456  459  460  461
462  463
## [166]  464  480  483  484  485  486  487  488  490  491  492  493  494
495  497
## [181]  498  499  500  501  502  504  550  552  560  576  600  603  604
605  606
## [196]  607  609  610  611  612  613  614  616  617  618  619  620  621
623  624
## [211]  648  674  695  696  697  699  700  701  702  703  706  707  708
709  710
## [226]  712  713  719  720  744  768  771  772  773  780  783  787  792
816  840
## [241]  864  888  912  936  960  984 1008 1009 1011 1012 1013 1014 1015
1016 1018
## [256] 1019 1020 1021 1022 1023 1025 1026 1027 1028 1029 1030 1032 1034
1035 1036
## [271] 1037 1038 1039 1041 1042 1043 1044 1045 1046 1047 1049 1050 1051
1052 1054
## [286] 1055 1056 1057 1059 1060 1061 1062 1063 1064 1066 1067 1068 1069
1070 1071
## [301] 1080 1104 1107 1108 1109 1110 1111 1112 1114 1115 1116 1117 1118
1119 1121
## [316] 1122 1123 1124 1125 1128 1152 1155 1161 1168 1176 1200 1210 1222
1224 1248
## [331] 1272 1274 1275 1276 1277 1278 1279 1281 1287 1289 1290 1291 1292
1294 1295
## [346] 1296 1298 1299 1300 1301 1302 1303 1305 1306 1307 1308 1309 1310
1311 1313
## [361] 1314 1315 1316 1318 1319 1320 1323 1324 1325 1326 1327 1329 1330
1331 1332
## [376] 1333 1334 1336 1337 1338 1339 1340 1341 1343 1344 1345 1347 1348
1349 1350
## [391] 1351 1352 1354 1355 1356 1357 1358 1359 1361 1362 1363 1364 1365
1366 1368
## [406] 1392 1416 1440 1462 1464 1474 1486 1488 1499 1500 1501 1502 1505
1507 1510
## [421] 1512 1536 1560 1584 1632 1649 1650 1651 1652 1653 1655 1656 1680
1704 1705
## [436] 1707 1708 1709 1710 1711 1712 1714 1715 1716 1717 1718 1719 1721
1722 1723
## [451] 1724 1725 1726 1728 1752 1776 1800 1824 1848 1872 1896 1897 1899
1900 1901
## [466] 1902 1903 1904 1906 1907 1908 1909 1910 1911 1913 1914 1915 1916
```

```
1917 1918
## [481] 1920 1944 1968 1969 1971 1972 1973 1974 1975 1976 1978 1979 1980
1981 1982
## [496] 1983 1985 1986 1987 1988 1989 1990 1992 1995 1996 1997 1998 1999
2001 2002
## [511] 2003 2004 2005 2006 2008 2009 2010 2011 2012 2013 2015 2016 2040
2041 2043
## [526] 2044 2045 2046 2047 2048 2050 2051 2052 2053 2054 2055 2057 2058
2059 2060
## [541] 2061 2062 2064 2065 2068 2069 2070 2071 2072 2073 2074 2075 2076
2077 2078
## [556] 2079 2083 2084 2085 2087 2088 2089 2091 2092 2093 2094 2095 2096
2098 2099
## [571] 2100 2101 2102 2103 2105 2106 2107 2108 2109 2110 2112 2113 2115
2116 2117
## [586] 2118 2119 2121 2122 2123 2124 2125 2126 2127 2129 2130 2131 2132
2133 2134
## [601] 2136 2146 2147 2148 2149 2150 2151 2153 2154 2155 2156 2157 2158
2160 2184
## [616] 2208 2232 2233 2235 2236 2237 2238 2239 2240 2242 2243 2244 2245
2246 2247
## [631] 2249 2250 2251 2252 2253 2254 2256 2280 2304 2306 2307 2308 2309
2310 2311
## [646] 2313 2314 2315 2316 2317 2318 2319 2321 2322 2323 2324 2326 2327
2328 2338
## [661] 2350 2352 2362 2374 2376 2400 2401 2403 2404 2405 2406 2407 2408
2410 2411
## [676] 2412 2413 2414 2415 2417 2418 2419 2420 2421 2422 2424 2448 2472
2489 2490
## [691] 2491 2492 2493 2494 2496 2520 2522 2546 2547 2548 2549 2550 2551
2552 2554
## [706] 2555 2556 2557 2558 2559 2561 2562 2563 2564 2566 2567 2568 2570
2571 2572
## [721] 2573 2574 2575 2577 2578 2579 2580 2581 2582 2583 2585 2586 2587
2588 2590
## [736] 2591 2592 2595 2596 2597 2598 2599 2601 2602 2603 2604 2605 2606
2608 2609
## [751] 2610 2611 2612 2613 2615 2616 2619 2620 2621 2622 2623 2625 2626
2627 2628
## [766] 2629 2630 2632 2633 2634 2635 2636 2637 2639 2640 2648 2659 2664
2688 2691
## [781] 2692 2693 2694 2695 2696 2698 2699 2700 2701 2702 2703 2712 2713
2715 2716
## [796] 2717 2718 2719 2720 2722 2723 2724 2725 2726 2727 2729 2730 2731
2732 2733
## [811] 2734 2736 2760 2784 2832 2840

which(is.na(data$Level))
```

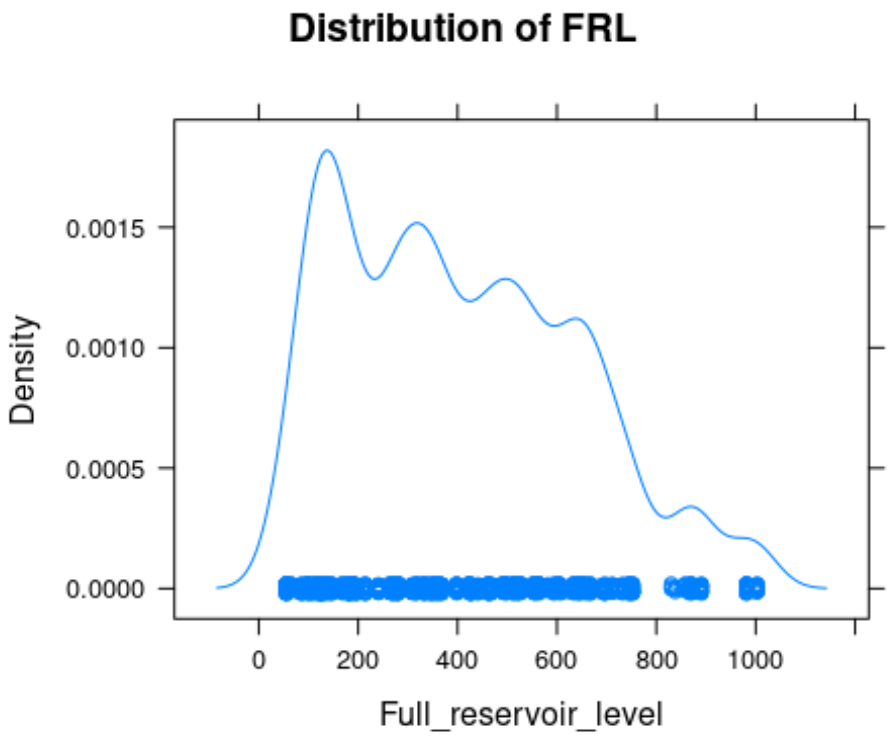
```
##   [1]   24   48   51   52   53   54   55   57   58   59   60   61   62
64   65
##  [16]   66   67   68   69   71   72   73   75   76   77   78   79   80
82   83
##  [31]   84   85   86   87   89   90   91   92   93   94   96  120  144
168  190
##  [46]  192  216  217  219  220  221  222  223  224  226  227  228  229
230  231
##  [61]  233  234  235  236  237  238  240  264  265  267  268  269  270
271  272
##  [76]  274  275  276  277  278  279  281  282  283  284  285  286  288
312  314
##  [91]  315  316  317  318  319  321  322  323  324  325  326  327  329
330  331
## [106]  332  334  335  336  338  339  340  341  342  343  345  346  347
348  349
## [121]  350  351  353  354  355  356  358  359  360  384  386  394  395
396  397
## [136]  398  399  408  409  432  433  435  436  437  438  439  440  442
443  444
## [151]  445  446  447  449  450  451  452  453  454  456  459  460  461
462  463
## [166]  464  480  483  484  485  486  487  488  490  491  492  493  494
495  497
## [181]  498  499  500  501  502  504  550  552  560  576  600  603  604
605  606
## [196]  607  609  610  611  612  613  614  616  617  618  619  620  621
623  624
## [211]  648  674  695  696  697  699  700  701  702  703  706  707  708
709  710
## [226]  712  713  719  720  744  768  771  772  773  780  783  787  792
816  840
## [241]  864  888  912  936  960  984 1008 1009 1011 1012 1013 1014 1015
1016 1018
## [256] 1019 1020 1021 1022 1023 1025 1026 1027 1028 1029 1030 1032 1034
1035 1036
## [271] 1037 1038 1039 1041 1042 1043 1044 1045 1046 1047 1049 1050 1051
1052 1054
## [286] 1055 1056 1057 1059 1060 1061 1062 1063 1064 1066 1067 1068 1069
1070 1071
## [301] 1080 1104 1107 1108 1109 1110 1111 1112 1114 1115 1116 1117 1118
1119 1121
## [316] 1122 1123 1124 1125 1128 1152 1155 1161 1168 1176 1200 1210 1222
1224 1248
## [331] 1272 1274 1275 1276 1277 1278 1279 1281 1287 1289 1290 1291 1292
1294 1295
## [346] 1296 1298 1299 1300 1301 1302 1303 1305 1306 1307 1308 1309 1310
1311 1313
## [361] 1314 1315 1316 1318 1319 1320 1323 1324 1325 1326 1327 1329 1330
1331 1332
```

```
## [376] 1333 1334 1336 1337 1338 1339 1340 1341 1343 1344 1345 1347 1348
1349 1350
## [391] 1351 1352 1354 1355 1356 1357 1358 1359 1361 1362 1363 1364 1365
1366 1368
## [406] 1392 1416 1440 1462 1464 1474 1486 1488 1499 1500 1501 1502 1505
1507 1510
## [421] 1512 1536 1560 1584 1632 1649 1650 1651 1652 1653 1655 1656 1680
1704 1705
## [436] 1707 1708 1709 1710 1711 1712 1714 1715 1716 1717 1718 1719 1721
1722 1723
## [451] 1724 1725 1726 1728 1752 1776 1800 1824 1848 1872 1896 1897 1899
1900 1901
## [466] 1902 1903 1904 1906 1907 1908 1909 1910 1911 1913 1914 1915 1916
1917 1918
## [481] 1920 1944 1968 1969 1971 1972 1973 1974 1975 1976 1978 1979 1980
1981 1982
## [496] 1983 1985 1986 1987 1988 1989 1990 1992 1995 1996 1997 1998 1999
2001 2002
## [511] 2003 2004 2005 2006 2008 2009 2010 2011 2012 2013 2015 2016 2040
2041 2043
## [526] 2044 2045 2046 2047 2048 2050 2051 2052 2053 2054 2055 2057 2058
2059 2060
## [541] 2061 2062 2064 2065 2068 2069 2070 2071 2072 2073 2074 2075 2076
2077 2078
## [556] 2079 2083 2084 2085 2087 2088 2089 2091 2092 2093 2094 2095 2096
2098 2099
## [571] 2100 2101 2102 2103 2105 2106 2107 2108 2109 2110 2112 2113 2115
2116 2117
## [586] 2118 2119 2121 2122 2123 2124 2125 2126 2127 2129 2130 2131 2132
2133 2134
## [601] 2136 2146 2147 2148 2149 2150 2151 2153 2154 2155 2156 2157 2158
2160 2184
## [616] 2208 2232 2233 2235 2236 2237 2238 2239 2240 2242 2243 2244 2245
2246 2247
## [631] 2249 2250 2251 2252 2253 2254 2256 2280 2304 2306 2307 2308 2309
2310 2311
## [646] 2313 2314 2315 2316 2317 2318 2319 2321 2322 2323 2324 2326 2327
2328 2338
## [661] 2350 2352 2362 2374 2376 2400 2401 2403 2404 2405 2406 2407 2408
2410 2411
## [676] 2412 2413 2414 2415 2417 2418 2419 2420 2421 2422 2424 2448 2472
2489 2490
## [691] 2491 2492 2493 2494 2496 2520 2522 2546 2547 2548 2549 2550 2551
2552 2554
## [706] 2555 2556 2557 2558 2559 2561 2562 2563 2564 2566 2567 2568 2570
2571 2572
## [721] 2573 2574 2575 2577 2578 2579 2580 2581 2582 2583 2585 2586 2587
2588 2590
## [736] 2591 2592 2595 2596 2597 2598 2599 2601 2602 2603 2604 2605 2606
2608 2609
```

```
## [751] 2610 2611 2612 2613 2615 2616 2619 2620 2621 2622 2623 2625 2626
2627 2628
## [766] 2629 2630 2632 2633 2634 2635 2636 2637 2639 2640 2648 2659 2664
2688 2691
## [781] 2692 2693 2694 2695 2696 2698 2699 2700 2701 2702 2703 2712 2713
2715 2716
## [796] 2717 2718 2719 2720 2722 2723 2724 2725 2726 2727 2729 2730 2731
2732 2733
## [811] 2734 2736 2760 2784 2832 2840

#Removing the NA values
data=na.omit(data)
summary(data)

##  Reservoir_name        Basin              Subbasin          Agency_name
##  Length:2040        Length:2040        Length:2040        Length:2040
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      Date                 Year          Month   Full_reservoir_level
##  Length:2040        Min.   :2023   Min.   :2   Min.   :  55.0
##  Class :character   1st Qu.:2023   1st Qu.:2   1st Qu.: 189.6
##  Mode  :character   Median :2023   Median :2   Median : 359.7
##                     Mean   :2023   Mean   :2   Mean   : 405.0
##                     3rd Qu.:2023   3rd Qu.:2   3rd Qu.: 564.0
##                     Max.   :2023   Max.   :2   Max.   :1002.8
##  Live_capacity_FRL    Storage            Level
##  Min.   :0.006     Min.   :0.00105   Min.   : 45.97
##  1st Qu.:0.300     1st Qu.:0.17603   1st Qu.:186.23
##  Median :0.697     Median :0.36735   Median :354.57
##  Mean   :1.311     Mean   :0.73986   Mean   :398.57
##  3rd Qu.:1.472     3rd Qu.:0.86922   3rd Qu.:550.35
##  Max.   :9.745     Max.   :6.35500   Max.   :972.19

#date type conversion
data["Date"]=as.Date(data$Date)
str(data)

## 'data.frame':    2040 obs. Of  11 variables:
##  $ Reservoir_name      : chr  "Aliyar Reservoir" "Aliyar Reservoir"
"Aliyar Reservoir" "Aliyar Reservoir" …
##  $ Basin               : chr  "West flowing rivers from Tadri to
Kanyakumari Basin" "West flowing rivers from Tadri to Kanyakumari Basin"
"West flowing rivers from Tadri to Kanyakumari Basin" "West flowing rivers
from Tadri to Kanyakumari Basin" …
##  $ Subbasin            : chr  "Varrar and others" "Varrar and others"
"Varrar and others" "Varrar and others" …
##  $ Agency_name         : chr  "CWC" "CWC" "CWC" "CWC" …
##  $ Date                : Date, format: "2023-02-01" "2023-02-02" …
```

```
##  $ Year                : int  2023 2023 2023 2023 2023 2023 2023 2023 2023
2023 …
##  $ Month               : int  2 2 2 2 2 2 2 2 2 …
##  $ Full_reservoir_level: num  320 320 320 320 320 …
##  $ Live_capacity_FRL   : num  0.095 0.095 0.095 0.095 0.095 0.095 0.095
0.095 0.095 0.095 …
##  $ Storage             : num  0.0579 0.0568 0.0559 0.0549 0.0537 …
##  $ Level               : num  314 314 314 313 313 …
##  - attr(*, "na.action")= 'omit' Named int [1:816] 24 48 51 52 53 54 55 57
58 59 …
##   ..- attr(*, "names")= chr [1:816] "24" "48" "51" "52" …
```

```
#Density plot to find the weightage of the Full_reservoir_level
densityplot(~Full_reservoir_level,data=data,main="Distribution of FRL")
```

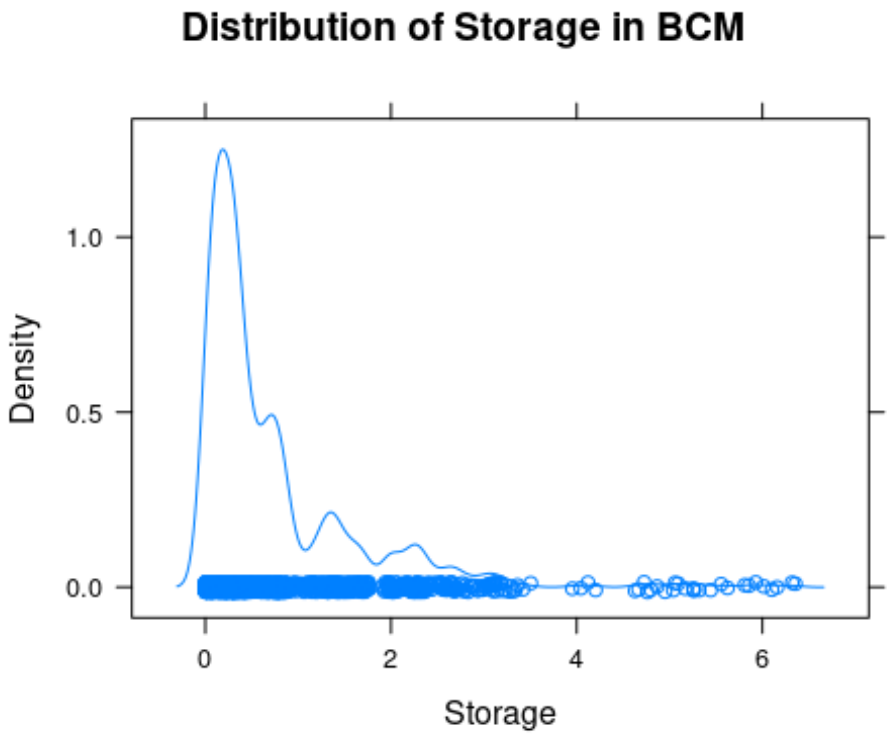



Inference:

The above graph shows that the Full_reservoir_level is Right skewed and the peak is attained between 0 to 200M followed by 200 to 400M. There is almost the strong distribution between 0 to 800M of Full_reservoir_level.

```
#Subsetting the dataset based on the full_reservoir_level
very_small=subset(data,Full_reservoir_level<250)
small=subset(data,Full_reservoir_level>250 & Full_reservoir_level<500)
large=subset(data,Full_reservoir_level >500 & Full_reservoir_level < 750)
```

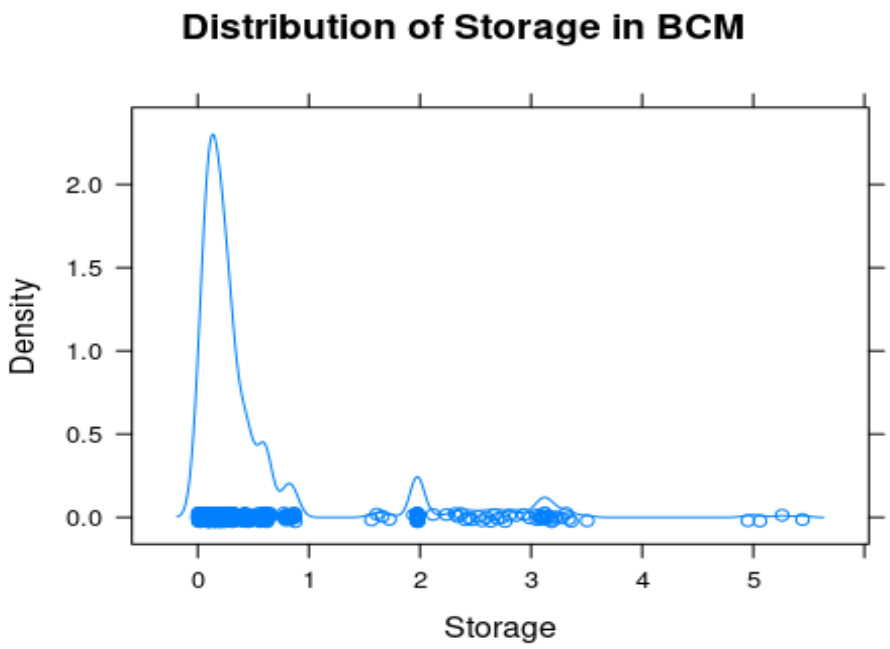
```
very_large=subset(data,Full_reservoir_level >750)

#Density plot to find  the Distribution of Storage in each subsets
densityplot(~Storage,data=data,main="Distribution of Storage in BCM")
```




Inference:

The above graph shows that the Storage of water (BCM) has more density in 0 to 3 BCM as for the whole India and the graph shows that it is a right skewed.

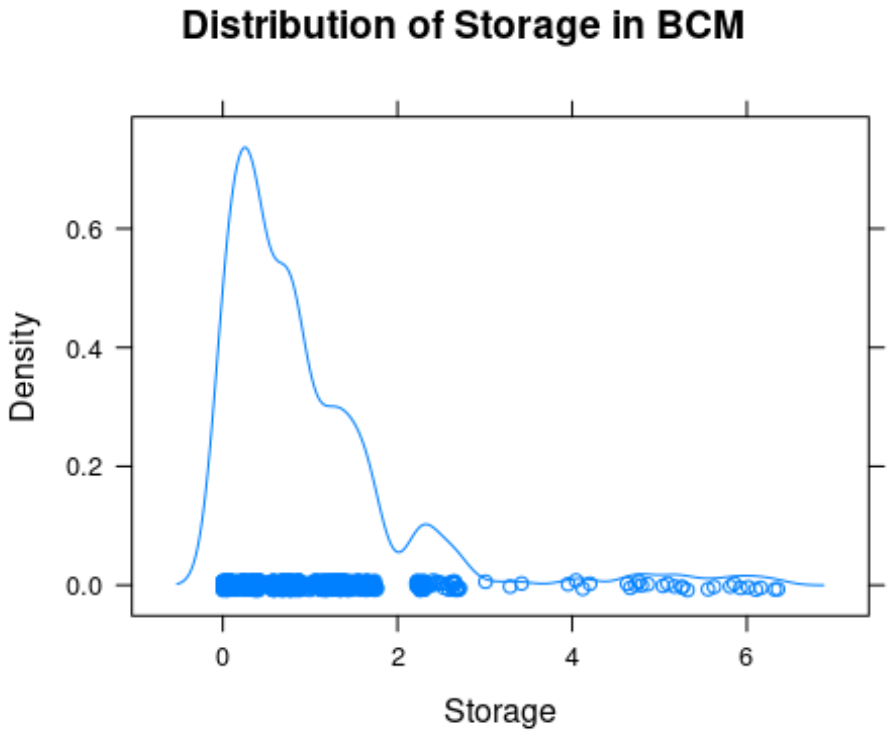
```
densityplot(~Storage,data=very_small,main="Distribution of Storage in BCM")
```




Inferences:

The Density plot shows that the data distribution is right skewed and most of the data points lie between 0 to 1 BCM of very small reservoirs in India.

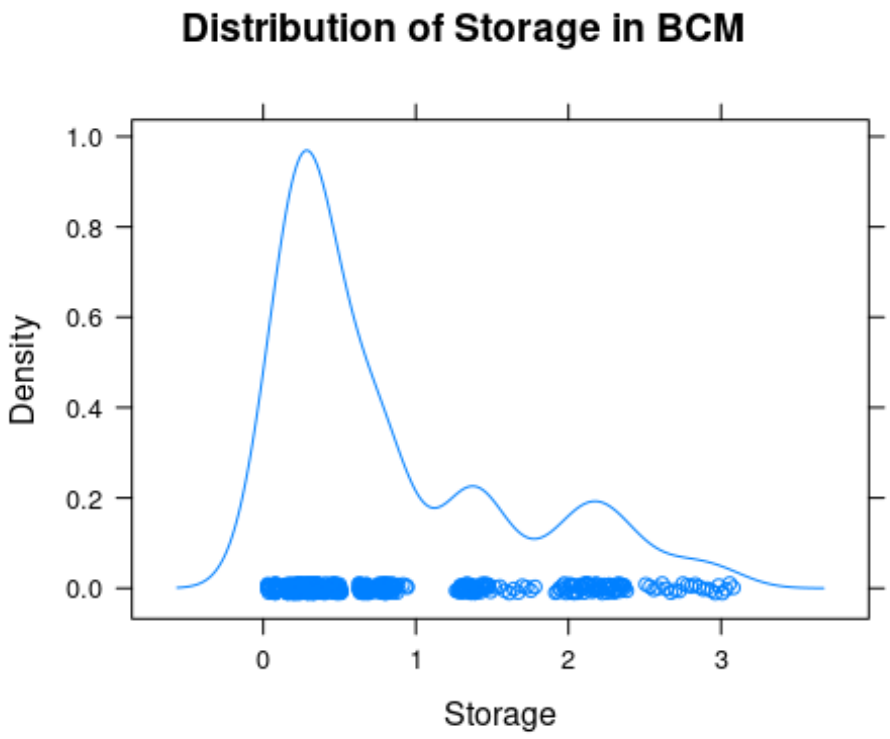
```
densityplot(~Storage,data=small,main="Distribution of Storage in BCM")
```



Inferences:

The above Density plot shows that the data distribution is right skewed and most of the data points lie between 0 to 2 BCM of water storage for small Reservoirs of India.

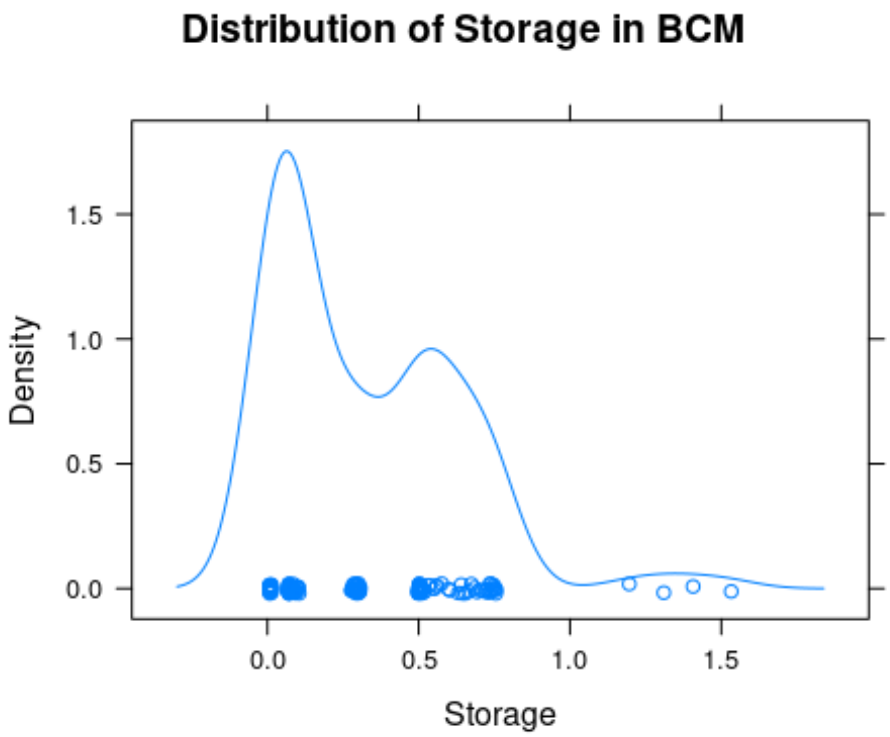
```
densityplot(~Storage,data=large,main="Distribution of Storage in BCM")
```




Inferences:

The above Density plot shows that the distribution of data is likely scattered and the highest thick density is occurred from 0 to 1 BCM of water Storage in Large Reservoirs of India.

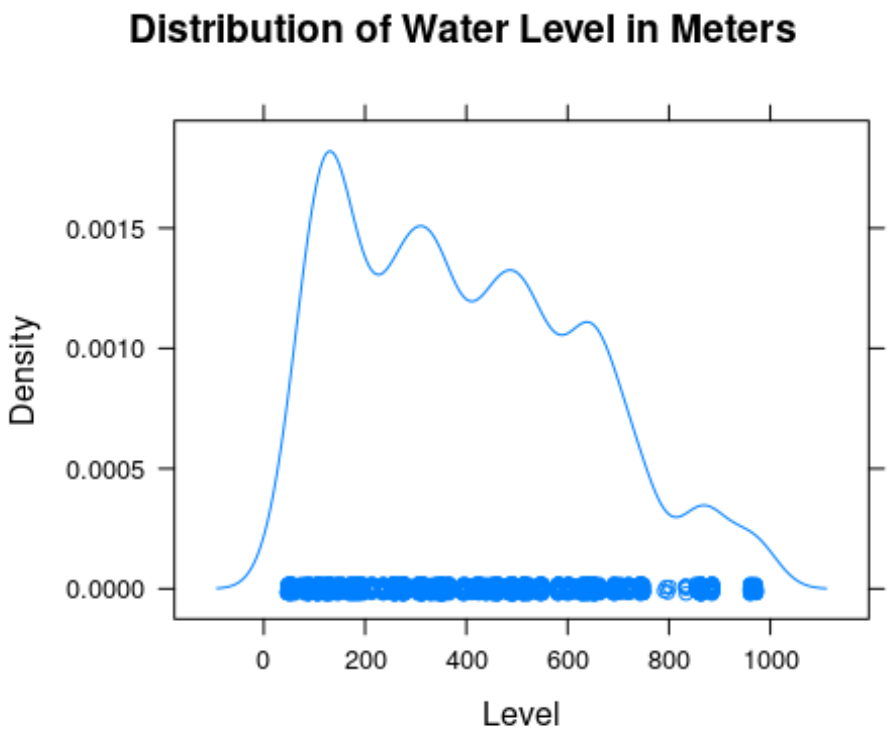
```
densityplot(~Storage,data=very_large,main="Distribution of Storage in BCM")
```




Inference:

The above density plot shows that the data distribution is right skewed and most of the points are lying from 0 to 1 BCM of storage and only 4 points are pointed greater than the 1BCM of water storage.

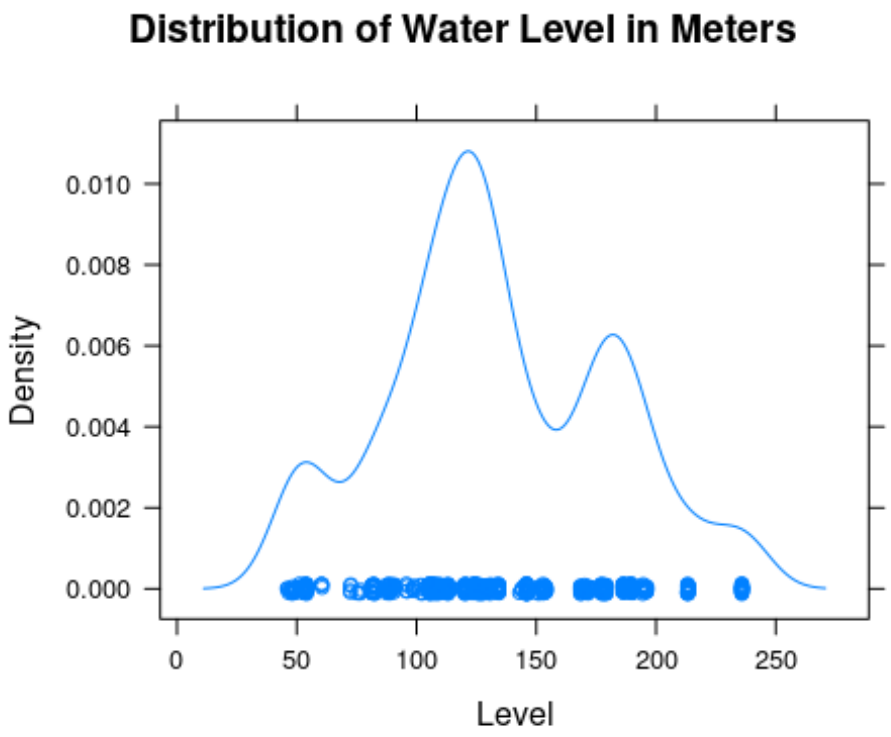
```
#Density plot to find the Distribution of water level in each subsets
densityplot(~Level,data=data,main="Distribution of Water Level in Meters")
```




Inferences:

The above density plot shows that the data distribution is right skewed and peaked is attained from 0 to 200M of the whole India Reservoirs.

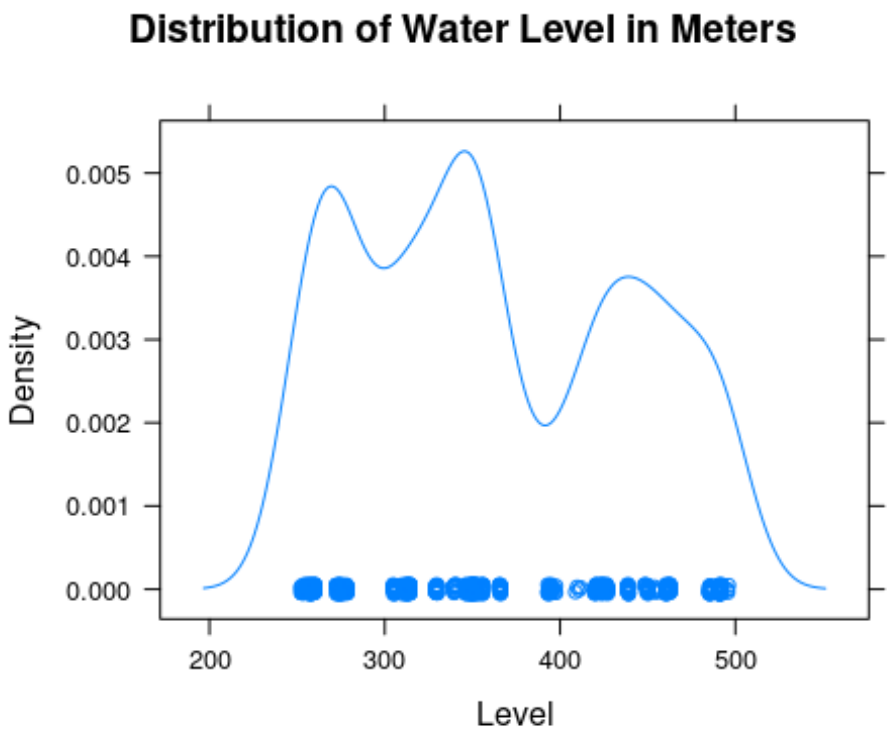
```
densityplot(~Level,data=very_small,main="Distribution of Water Level in
Meters")
```




Inference:

The above distribution of very_small reservoirs data is partially normally distributed. The highest peak is attained between 100 to 150M.

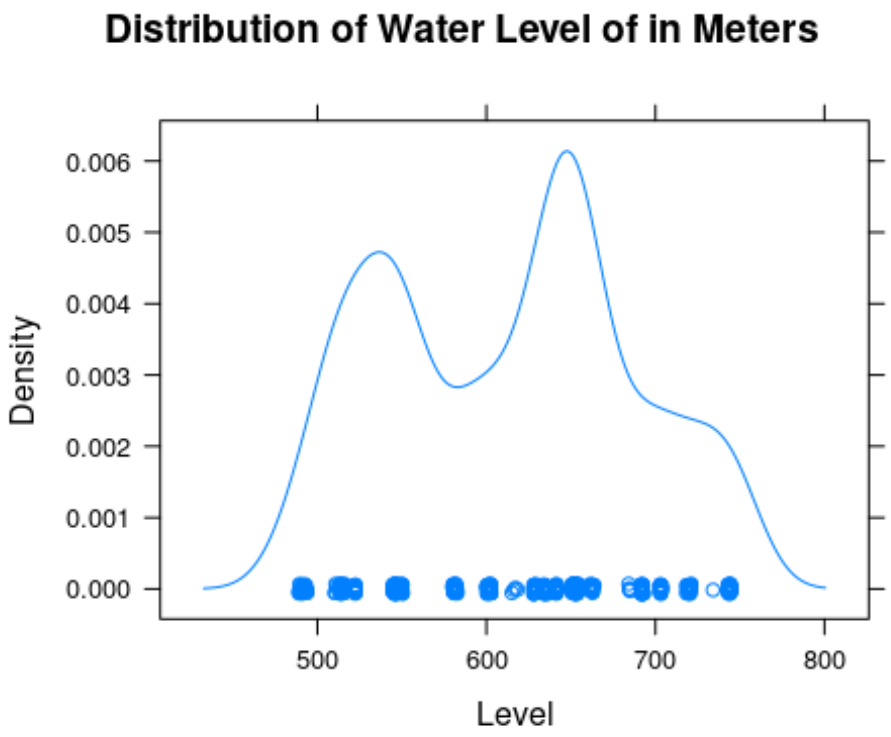
```
densityplot(~Level,data=small,main="Distribution of Water Level in Meters")
```




Inference:

The above density plot is partially normally distributed and their data points are lie between 250 to 500M of water Level.

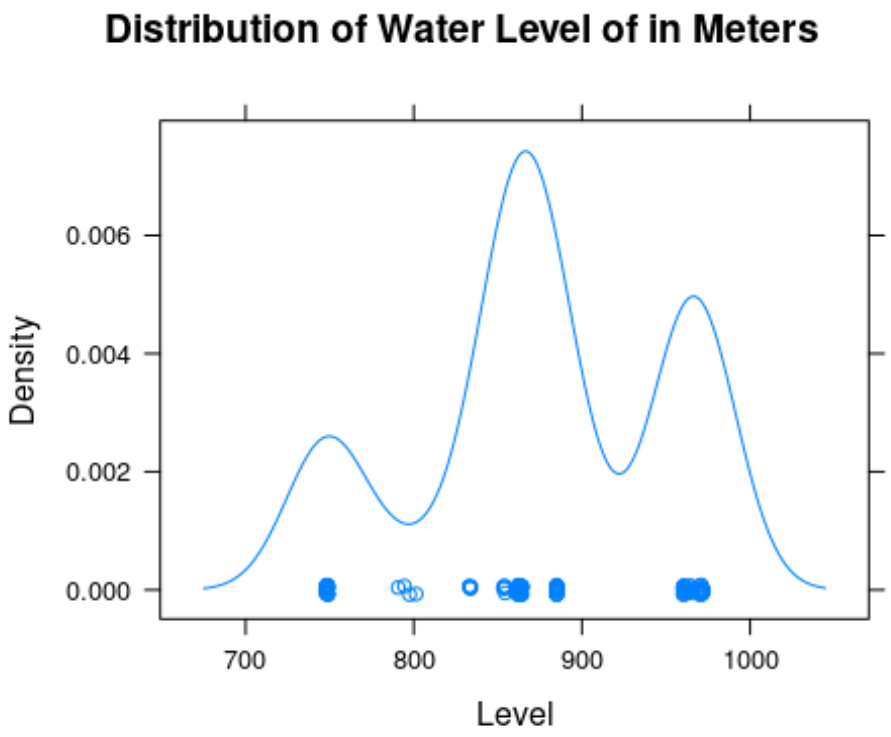
```
densityplot(~Level,data=large,main="Distribution of Water Level of in
Meters")
```




Inferences:

The above density plot is left skewed that the kurtosis is peaked at 600 to 700M of water Level.

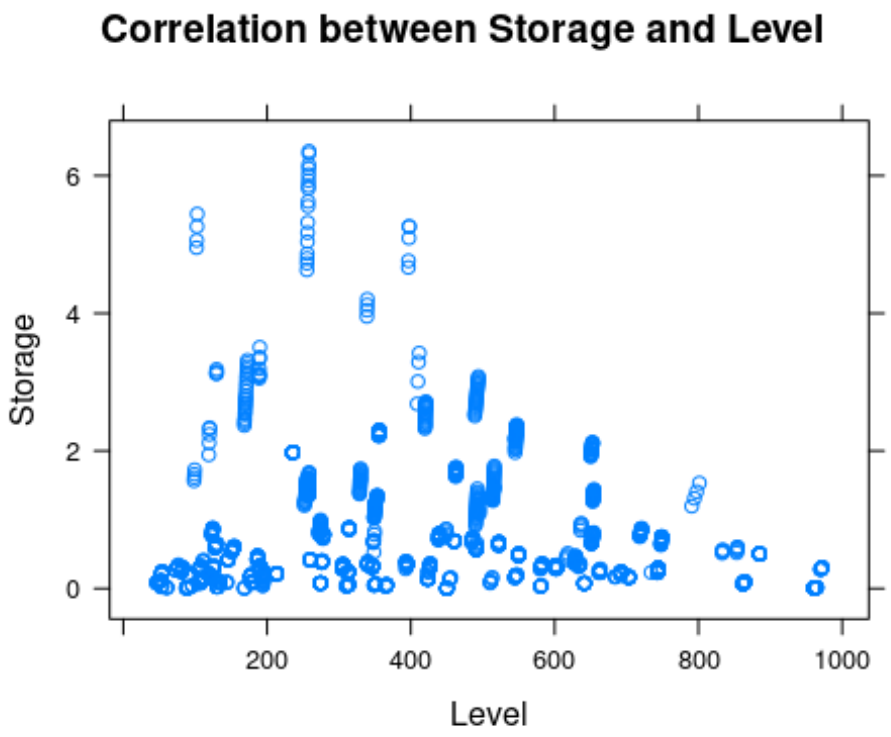
```
densityplot(~Level,data=very_large,main="Distribution of Water Level of in
Meters")
```




Inference:

The above density plot shows that the data distribution of very_large reservoirs water Level is right skewed and the data points are mostly scattered and it not forms any cluster.

```
#Scatter plot to find the correlation between Level and Storage
cor(data$Storage,data$Level)

## [1] -0.01680621

#Scattter plot for Storage and Level for the whole dataset
xyplot(Storage~Level,data=data,main="Correlation between Storage and Level")
```



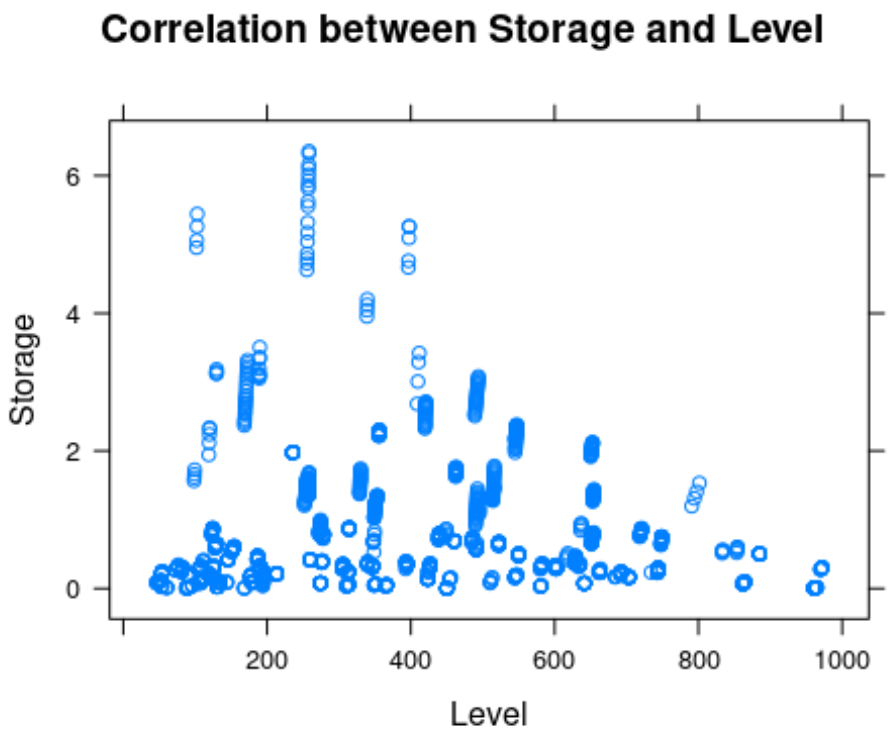

Inference:

There is a negative correlation between Storage and Level of whole dataset, it indicates that the water in the reservoir may be affected by some external factors such as water evaporation, drought etc., and correlation value is -0.01

```
cor(very_small$Storage,very_small$Level)
```

```
## [1] 0.3110995
```

```
#Scatter plot for Storage and Level for the very_small Reservoirs
xyplot(Storage~Level,data=data,main="Correlation between Storage and Level")
```



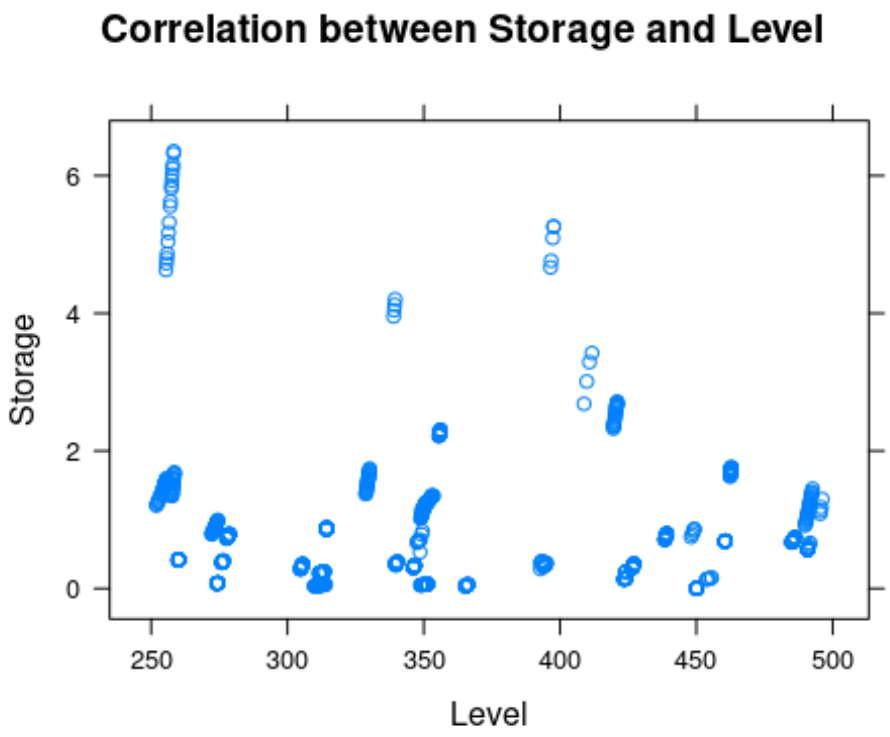

Inference:

There is a low positive correlation between Storage and Level of very_small reservoirs, it indicates that the water in the reservoir may be affected by some external factors such as water evaporation, drought etc., and their correlation value is 0.31

```
cor(small$Storage,small$Level)
```

```
## [1] -0.09551172
```

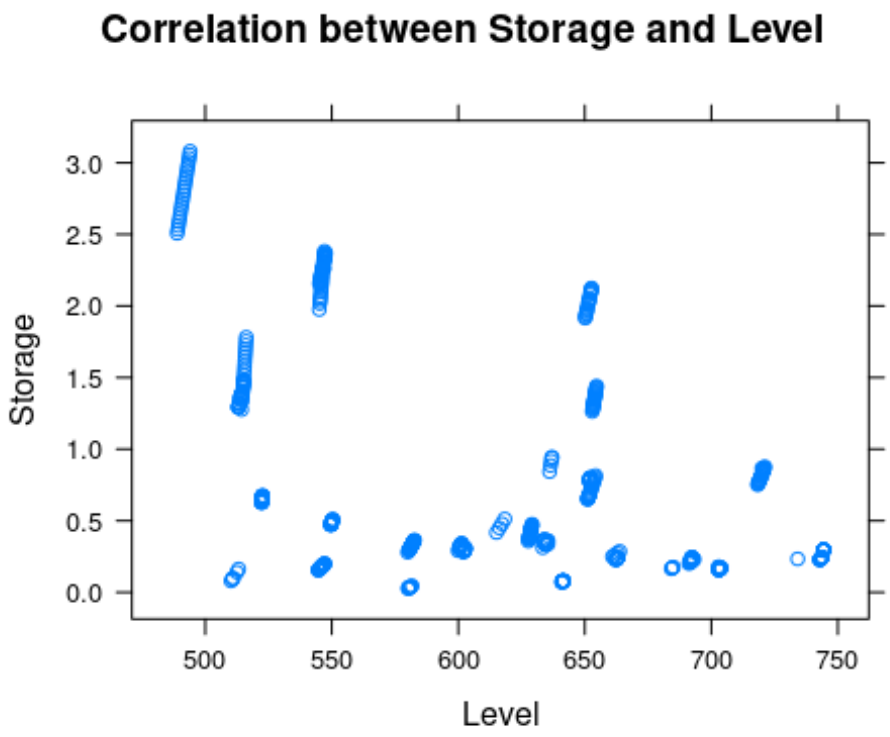
```
#Scatter plot for Storage and Level for the small Reservoirs
xyplot(Storage~Level,data=small,main="Correlation between Storage and Level")
```




Inference:

There is a negative correlation between Storage and Level of small reservoirs, it indicates that the water in the reservoir may be affected by some external factors such as water evaporation, drought etc., and correlation value is -0.09

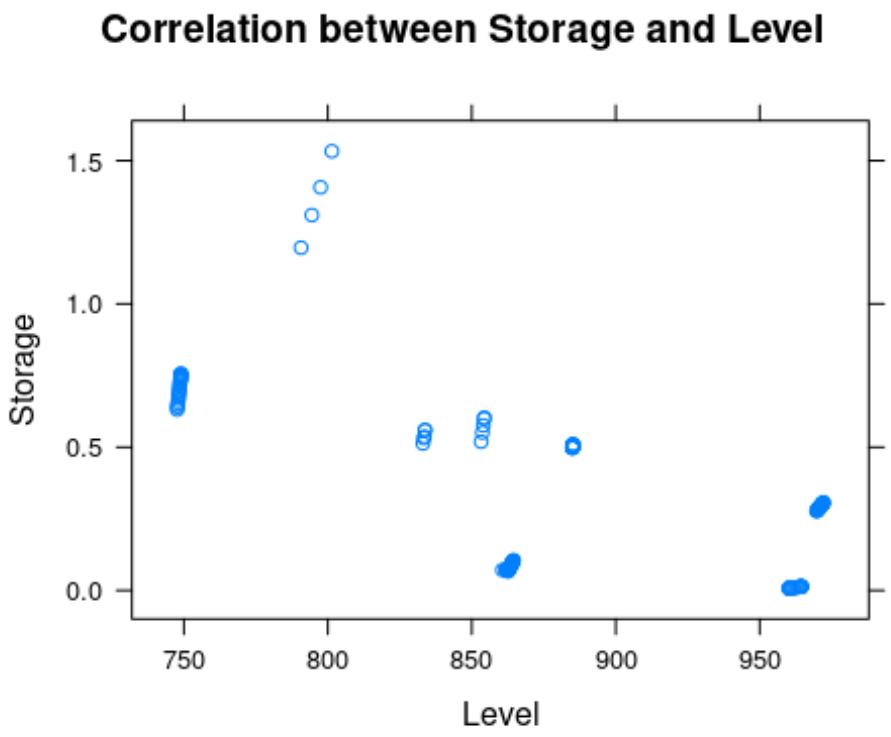
```
cor(large$Storage,large$Level)

## [1] -0.4706887

#Scatter plot for Storage and Level for the large Reservoirs
xyplot(Storage~Level,data=large,main="Correlation between Storage and Level")
```




Inference:

There is a negative correlation between Storage and Level of large reservoirs, it indicates that the water in the reservoir may be affected by some external factors such as water evaporation, drought etc., and correlation value is -0.47

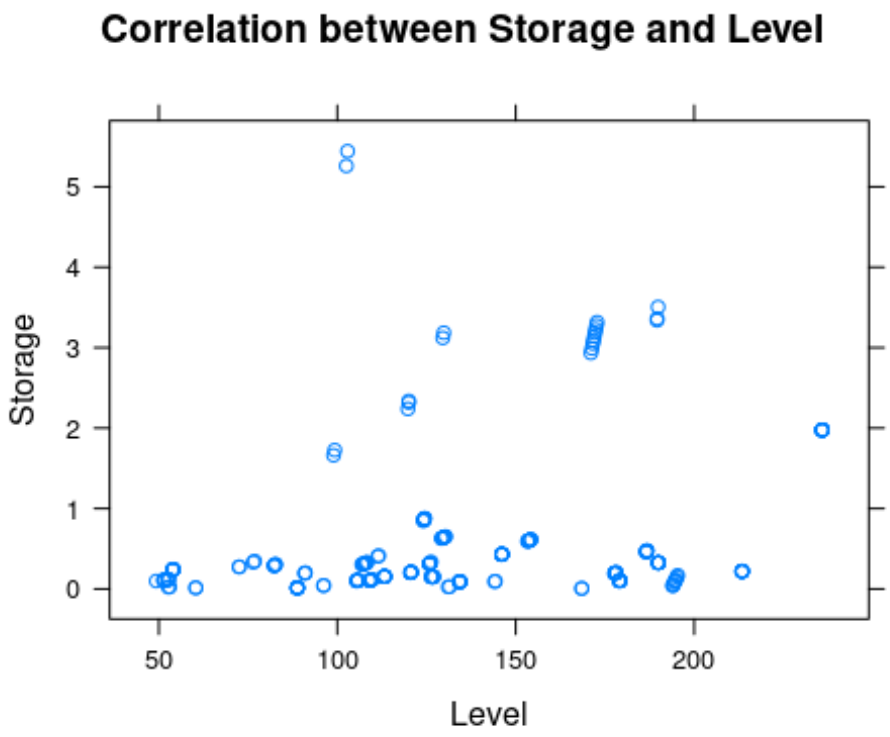
```
cor(very_large$Storage,very_large$Level)

## [1] -0.6080285

#Scatter plot for Storage and Level for the very_large Reservoirs
xyplot(Storage~Level,data=very_large,main="Correlation between Storage and
Level")
```




Inference:

There is a negative correlation between Storage and Level of very_large reservoirs, it indicates that the water in the reservoir may be affected by some external factors such as water evaporation, drought etc., and correlation value is -0.60

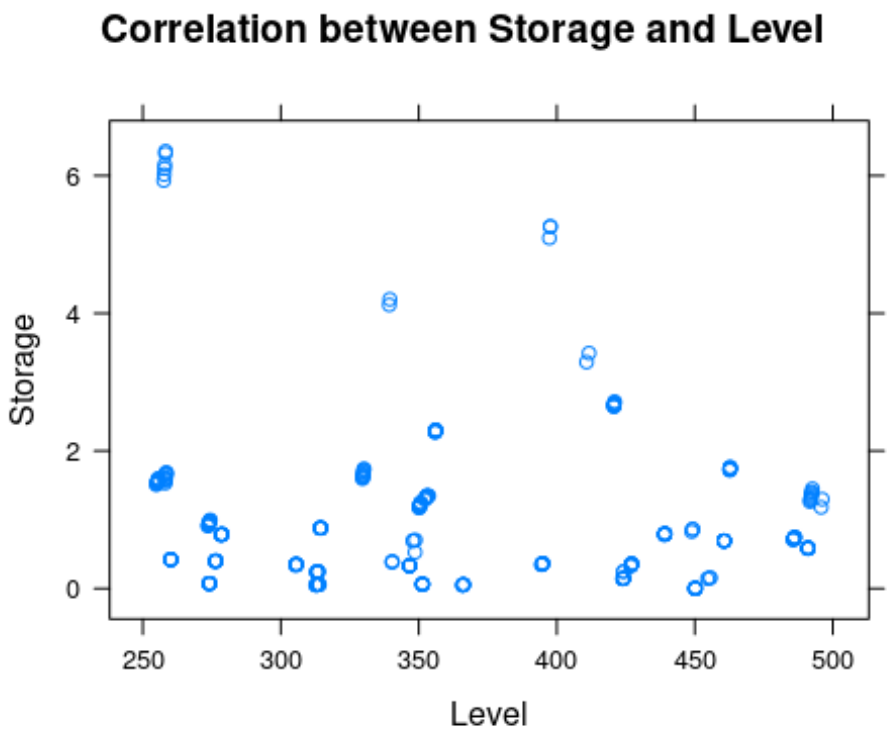
```
#Subsetting the dataset based on Date
sub1=subset(data,Date<"2023-02-10")
sub2=subset(data,Date>"2023-02-09" & Date<"2023-02-17")
sub3=subset(data,Date>"2023-02-16")

#Sub1 under very_small Storage category
sub1_very_small=subset(sub1,Full_reservoir_level<250)
cor(sub1_very_small$Storage,sub1_very_small$Level)

## [1] 0.2689758

xyplot(Storage~Level,data=sub1_very_small,main="Correlation between Storage
and Level")
```




Inference:

There is a very low positive correlation between the Storage and Level of first 9 days data of very_small reservoirs of India. This may be due to drought or rainfall etc., and the correlation value is 0.26.

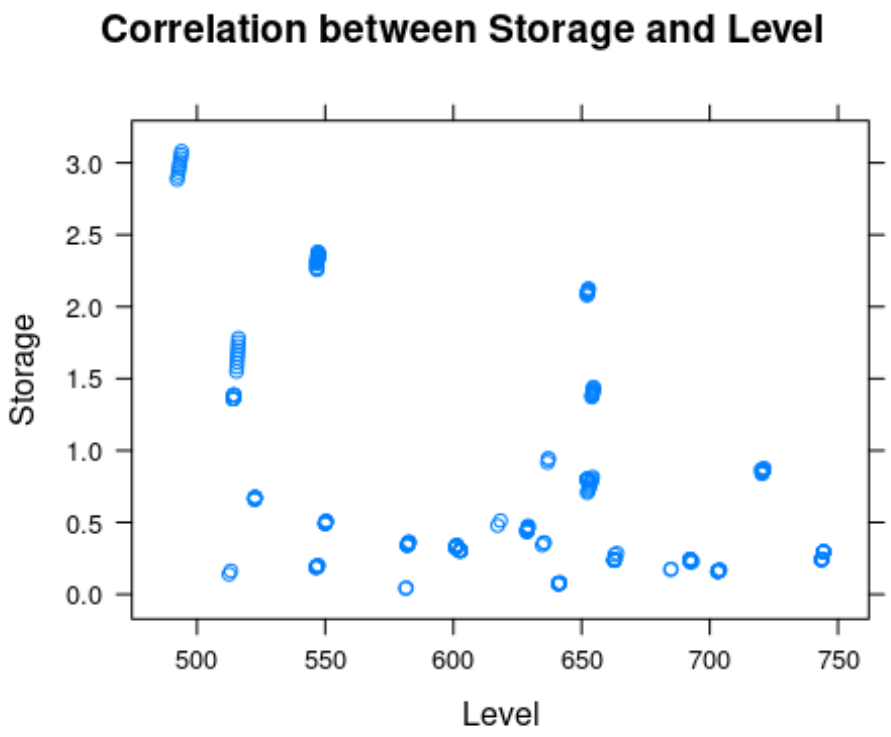
```
#Sub1 under small Storage category
sub1_small=subset(sub1,Full_reservoir_level>250 & Full_reservoir_level<500)
cor(sub1_small$Storage,sub1_small$Level)

## [1] -0.08981478

xyplot(Storage~Level,data=sub1_small,main="Correlation between Storage and
Level")
```




Inference:

There is a very low negative correlation between the Storage and Level of first 9 days data of small reservoirs of India. This may be due to drought or rainfall etc., and the correlation value is -0.08.

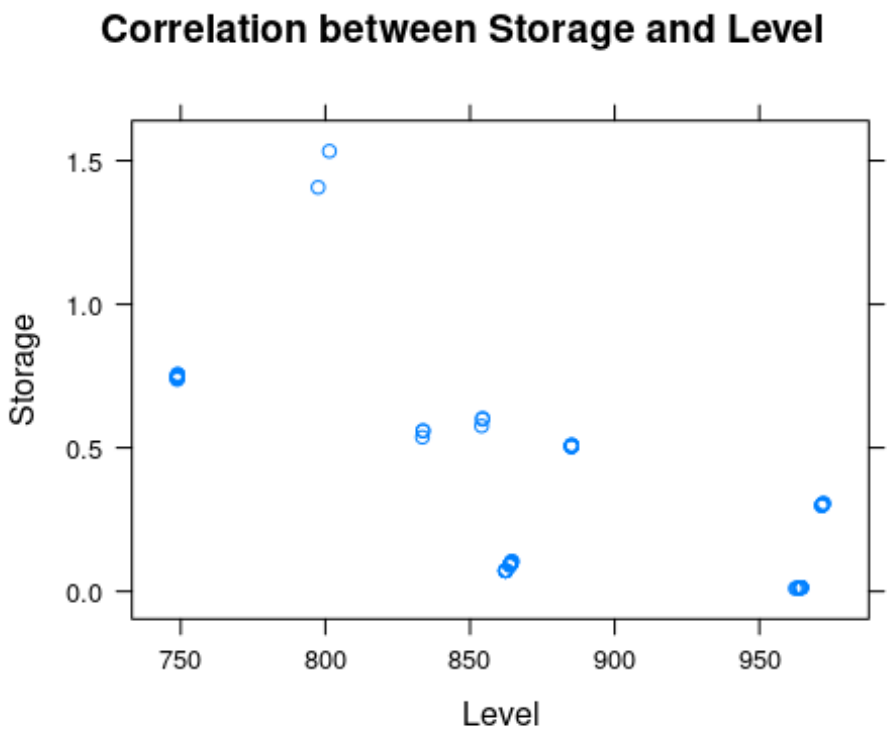
```
#Sub1 under Large Storage category
sub1_large=subset(sub1,Full_reservoir_level >500 & Full_reservoir_level <
750)
cor(sub1_large$Storage,sub1_large$Level)

## [1] -0.5041988

xyplot(Storage~Level,data=sub1_large,main="Correlation between Storage and
Level")
```




Inference:

There is a negative correlation between the Storage and Level of first 9 days data of large reservoirs of India. This may be due to drought or rainfall etc., and the correlation value is -0.50.

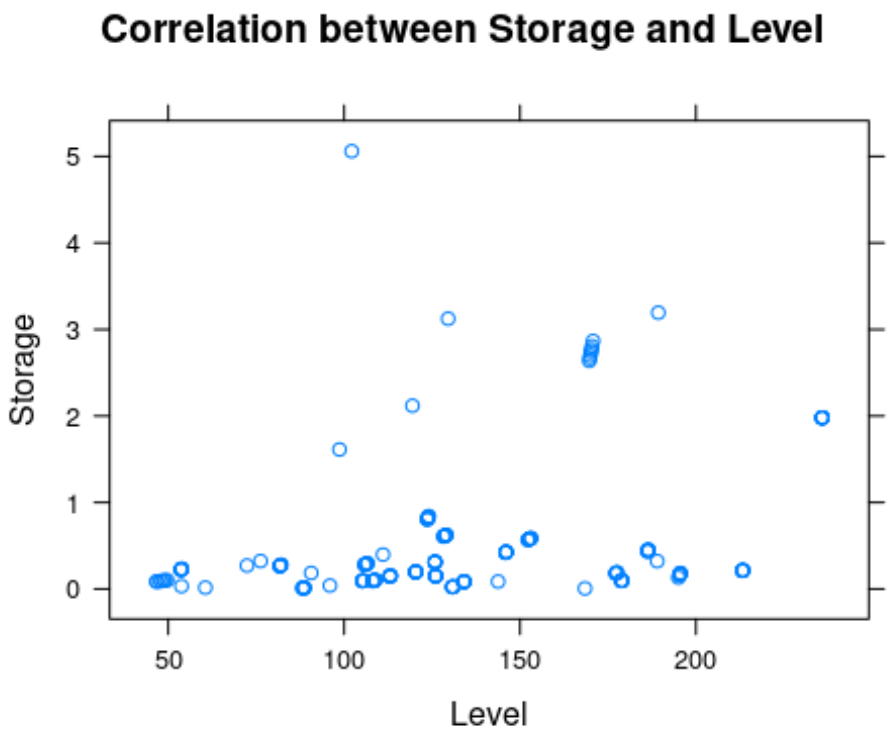
```
#Sub1 under Very_large Storage category
sub1_very_large=subset(sub1,Full_reservoir_level >750)
cor(sub1_very_large$Storage,sub1_very_large$Level)

## [1] -0.6145204

xyplot(Storage~Level,data=sub1_very_large,main="Correlation between Storage
and Level")
```




Inference:

There is a very low negative correlation between the Storage and Level of first 9 days data of very_large reservoirs of India. This may be due to drought or rainfall etc., and the correlation value is -0.61.

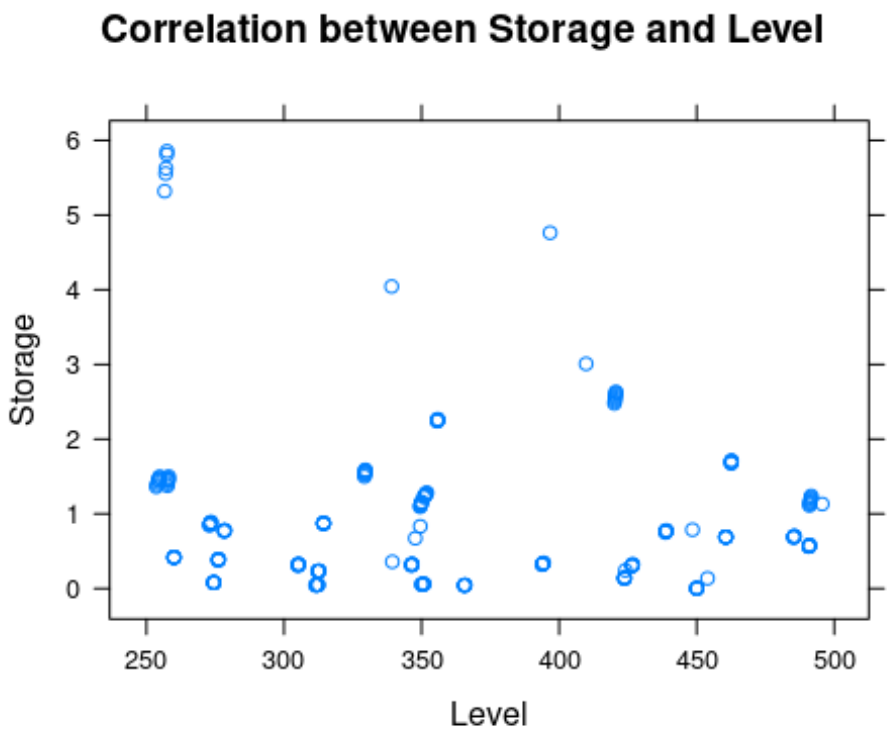
```
#Sub2 under very_small Storage category
sub2_very_small=subset(sub2,Full_reservoir_level<250)
cor(sub2_very_small$Storage,sub2_very_small$Level)

## [1] 0.3172206

xyplot(Storage~Level,data=sub2_very_small,main="Correlation between Storage
and Level")
```




Inference:

There is a low positive correlation between the Storage and Level of second 9 days data of very_small reservoirs of India. This may be due to drought or rainfall etc., and the correlation value is -0.31.

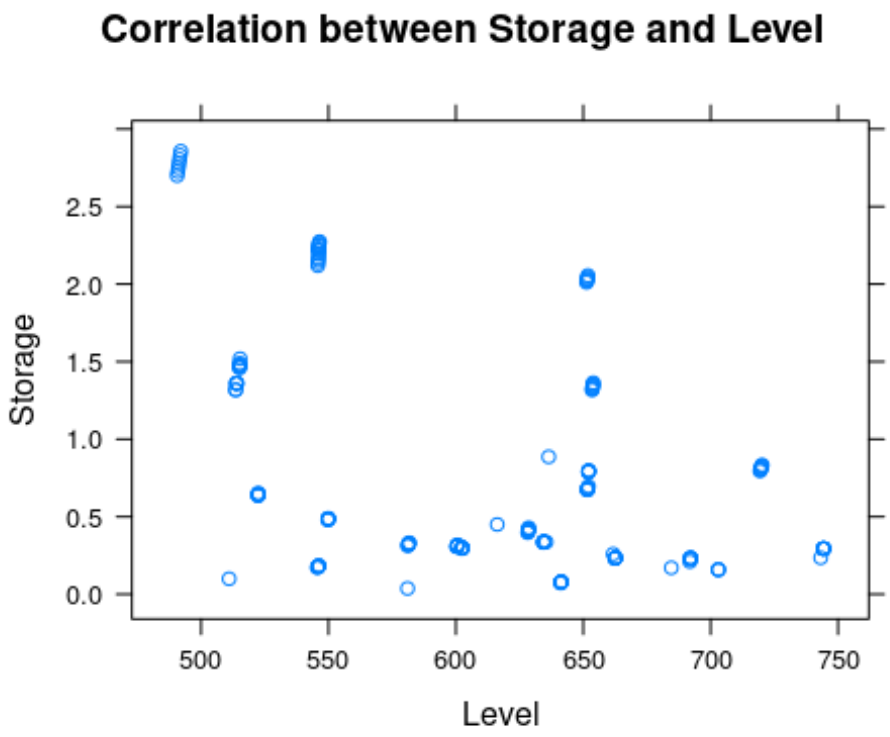
```
#Sub2 under very_small Storage category
sub2_small=subset(sub2,Full_reservoir_level>250 & Full_reservoir_level<500)
cor(sub2_small$Storage,sub2_small$Level)

## [1] -0.09926354

xyplot(Storage~Level,data=sub2_small,main="Correlation between Storage and
Level")
```




Inference:

There is a low negative correlation between the Storage and Level of second 9 days data of small reservoirs of India. This may be due to drought or rainfall etc., and the correlation value is -0.09.

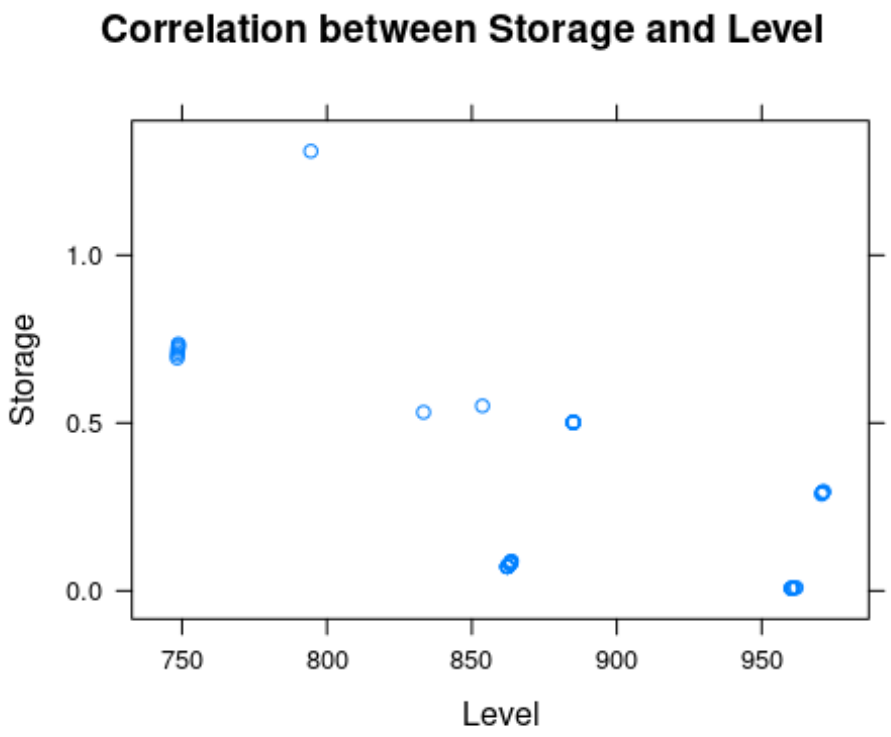
```
#Sub2 under large Storage category
sub2_large=subset(sub2,Full_reservoir_level >500 & Full_reservoir_level <
750)
cor(sub2_large$Storage,sub2_large$Level)

## [1] -0.4699136

xyplot(Storage~Level,data=sub2_large,main="Correlation between Storage and
Level")
```




Inference:

There is a negative correlation between the Storage and Level of second 9 days data of large reservoirs of India. This may be due to drought or rainfall etc., and the correlation value is -0.46.

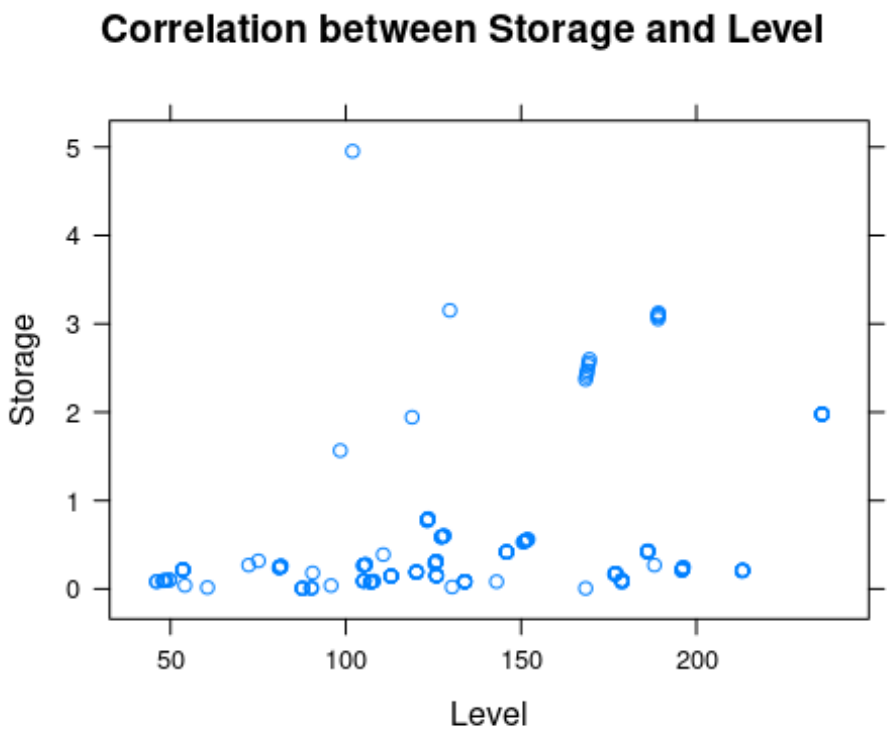
```
#Sub2 under very_large Storage category
sub2_very_large=subset(sub2,Full_reservoir_level >750)
cor(sub2_very_large$Storage,sub2_very_large$Level)

## [1] -0.61504

xyplot(Storage~Level,data=sub2_very_large,main="Correlation between Storage
and Level")
```




Inference:

There is a negative correlation between the Storage and Level of second 9 days data of very_large reservoirs of India. This may be due to drought or rainfall etc., and the correlation value is -0.61.

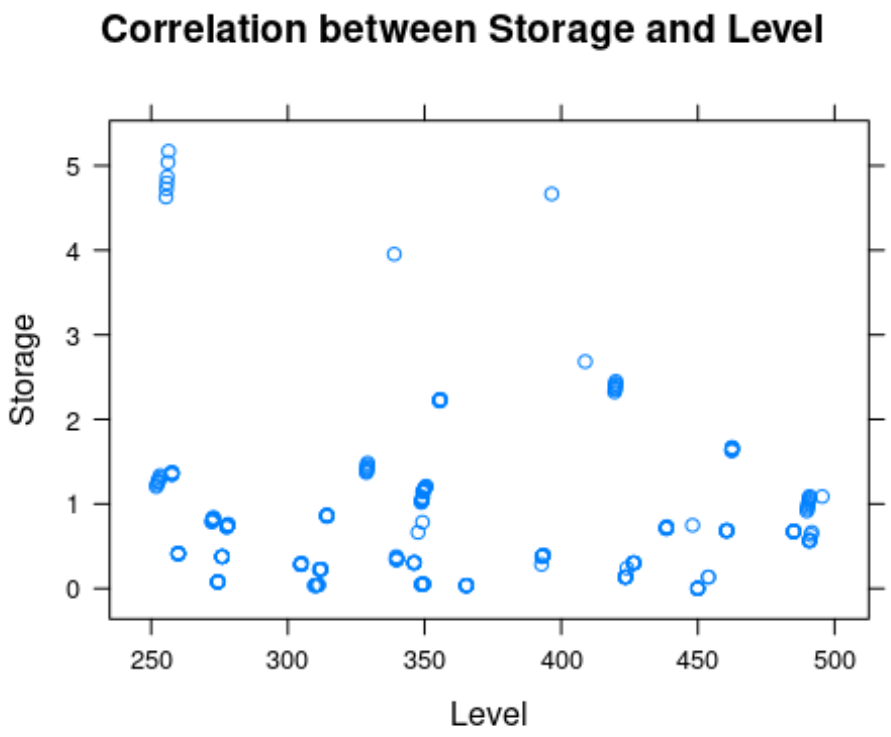
```
#Sub3 under very_small Storage category
sub3_very_small=subset(sub3,Full_reservoir_level<250)
cor(sub3_very_small$Storage,sub3_very_small$Level)

## [1] 0.3693982

xyplot(Storage~Level,data=sub3_very_small,main="Correlation between Storage
and Level")
```




Inference:

There is a low positive correlation between the Storage and Level of third 9 days data of very_small reservoirs of India. This may be due to drought or rainfall etc., and the correlation value is 0.36.

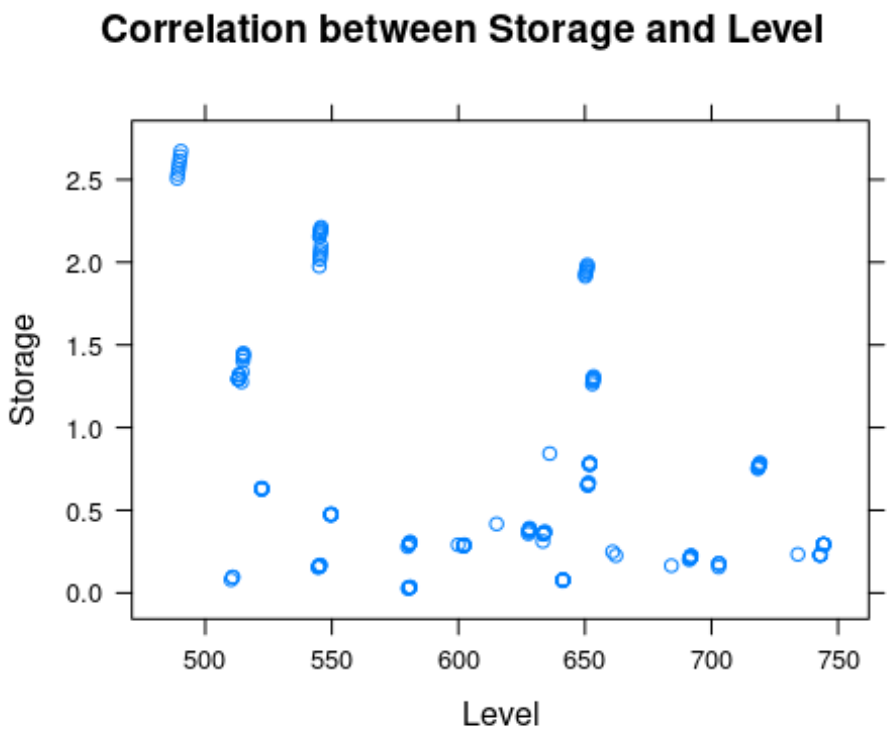
```
#Sub3 under small Storage category
sub3_small=subset(sub3,Full_reservoir_level>250 & Full_reservoir_level<500)
cor(sub3_small$Storage,sub3_small$Level)

## [1] -0.1046127

xyplot(Storage~Level,data=sub3_small,main="Correlation between Storage and
Level")
```




Inference:

There is a low negative correlation between the Storage and Level of third 9 days data of small reservoirs of India. This may be due to drought or rainfall etc., and the correlation value is -0.10.

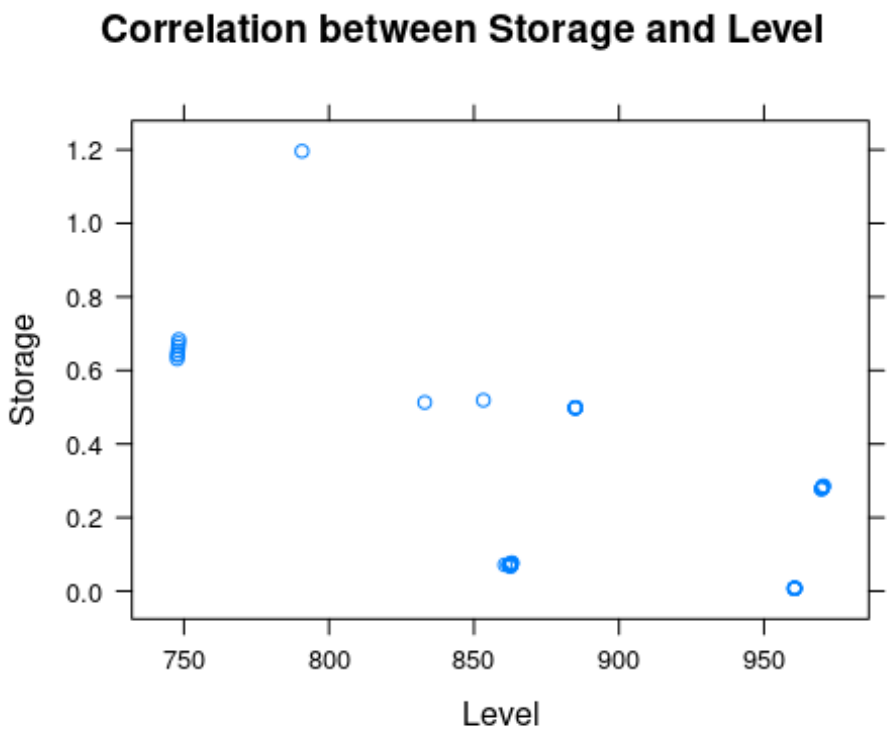
```
#Sub3 under small Storage category
sub3_large=subset(sub3,Full_reservoir_level >500 & Full_reservoir_level <
750)
cor(sub3_large$Storage,sub3_large$Level)

## [1] -0.4298773

xyplot(Storage~Level,data=sub3_large,main="Correlation between Storage and
Level")
```




Inference:

There is a negative correlation between the Storage and Level of third 9 days data of large reservoirs of India. This may be due to drought or rainfall etc., and the correlation value is -0.42.

```
#Sub3 under very_large Storage category
sub3_very_large=subset(sub3,Full_reservoir_level >750)
cor(sub3_very_large$Storage,sub3_very_large$Level)

## [1] -0.6034687

xyplot(Storage~Level,data=sub3_very_large,main="Correlation between Storage
and Level")
```




Inference:

There is a negative correlation between the Storage and Level of third 9 days data of very_large reservoirs of India. This may be due to drought or rainfall etc., and the correlation value is -0.60.

Insights:

The data set contains a data of water level of Indian reservoirs from 01/02/2023 to 24/02/23.

As a result of Exploratory Analysis, it is founded that there is no overall correlation between the Storage and Level of the water reservoirs.

The data of Reservoirs are subseted into four categories based on the Full_reservoir_level of water reservoir,

- Very_small (0M to 250M) of FRL
- Small (251M to 500M) of FRL
- Large (501M to 750M) of FRL
- Very_Large (above 751M) of FRL

This categorization of Reservoirs is very useful to how large Reservoirs are affected by external factors.,

After doing Univariate analysis,

- It is founded that **most of the Reservoirs have 0 to 3BCM** of water in it. It may vary according to the different category of Reservoirs.,
- It is founded that the **most of the Reservoirs have 0 to 200M of water Level** followed by 400M, 600M and 800M.
- **Only few** reservoirs have a water level **above 800M**

The Result of Bivariate analysis,

- The overall correlation between the Storage and Level results a negative. It indicates that the **water resources may be affected by any natural or artificial factors** such as Drought, Heavy rainfall, Evaporation of water etc.,
- **Very_small** Reservoirs have a positive correlation of all subsets of Date attribute and as a whole dataset. It indicates that it **does not affect by those factors so much.**
- Other three Reservoirs **Small, Large, Very_large** results a negative correlation which indicates that these reservoirs are **highly affected by those natural or artificial factors**.
- **Very_large Category** Reservoirs are **highly negatively correlated** compared to other categories
- It is a valid point to recommend that **Very_Large category reservoirs needs more attention to preserve the water resource** to prevent the loss of resource and to make it to available in the summer season.