**loading the dataset**

In [1]: 
```
data = read.csv("StatewiseTestingDetails.csv")
```

In [2]: `data #output`

A data.frame: 201 × 5

| Date | State | TotalSamples | Negative | Positive |
|------|-------|--------------|----------|----------|
| <fct> | <fct> | <int> | <int> | <int> |
| 2/17/2020 | Kerala | 423 | 406 | 3 |
| 2/18/2020 | Kerala | 425 | 420 | 3 |
| 2/19/2020 | Kerala | 432 | 423 | 3 |
| 2/20/2020 | Kerala | 433 | 423 | 3 |
| 2/21/2020 | Kerala | 437 | 426 | 3 |
| 2/22/2020 | Kerala | 441 | 436 | 3 |
| 2/23/2020 | Kerala | 444 | 436 | 3 |
| 2/24/2020 | Kerala | 449 | 441 | 3 |
| 2/25/2020 | Kerala | 451 | 446 | 3 |
| 2/26/2020 | Kerala | 460 | 451 | 3 |
| 2/27/2020 | Kerala | 462 | 451 | 3 |
| 2/28/2020 | Kerala | 471 | 463 | 3 |
| 2/29/2020 | Kerala | 485 | 471 | 3 |
| 3/1/2020 | Kerala | 488 | 471 | 3 |
| 3/1/2020 | Tamil Nadu | 49 | 49 | 0 |
| 3/2/2020 | Kerala | 504 | 493 | 3 |
| 3/2/2020 | Tamil Nadu | 50 | 50 | 0 |
| 3/3/2020 | Kerala | 520 | 494 | 3 |
| 3/3/2020 | Tamil Nadu | 50 | 50 | 0 |
| 3/4/2020 | Kerala | 552 | 511 | 3 |
| 3/4/2020 | Tamil Nadu | 54 | 54 | 0 |
| 3/5/2020 | Kerala | 574 | 535 | 3 |
| 3/5/2020 | Tamil Nadu | 54 | 54 | 0 |
| 3/6/2020 | Kerala | 631 | 579 | 3 |
| 3/6/2020 | Tamil Nadu | 56 | 54 | 0 |
| 3/7/2020 | Kerala | 682 | 616 | 3 |
| 3/7/2020 | Tamil Nadu | 60 | 56 | 1 |
| 3/8/2020 | Kerala | 729 | 664 | 8 |
| 3/8/2020 | Tamil Nadu | 68 | 59 | 1 |
| 3/9/2020 | Kerala | 807 | 717 | 9 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 4/8/2020 | Odisha | 2441 | 2399 | 42 |
| 4/8/2020 | Punjab | 2937 | 2614 | 106 |
| 4/8/2020 | Rajasthan | 17638 | 16401 | 363 |

| Date | State | TotalSamples | Negative | Positive |
|------|-------|--------------|----------|----------|
| <fct> | <fct> | <int> | <int> | <int> |
| 4/8/2020 | Tamil Nadu | 6095 | 4893 | 738 |
| 4/9/2020 | Andhra Pradesh | 5960 | 5597 | 363 |
| 4/9/2020 | Chandigarh | 199 | 169 | 18 |
| 4/9/2020 | Delhi | 9968 | 8643 | 720 |
| 4/9/2020 | Gujarat | 6199 | 5579 | 262 |
| 4/9/2020 | Jammu and Kashmir | 2649 | 2465 | 184 |
| 4/9/2020 | Karnataka | 7613 | 7176 | 197 |
| 4/9/2020 | Kerala | 12710 | 11469 | 357 |
| 4/9/2020 | Madhya Pradesh | 5135 | 3989 | 411 |
| 4/9/2020 | Odisha | 2841 | 2797 | 44 |
| 4/9/2020 | Punjab | 3192 | 2777 | 130 |
| 4/9/2020 | Rajasthan | 19107 | 17851 | 430 |
| 4/9/2020 | Tamil Nadu | 7267 | 5948 | 834 |
| 4/10/2020 | Andhra Pradesh | 6374 | 5993 | 381 |
| 4/10/2020 | Chandigarh | 223 | 199 | 19 |
| 4/10/2020 | Delhi | 11061 | 9662 | 903 |
| 4/10/2020 | Gujarat | 7718 | 7237 | 378 |
| 4/10/2020 | Jammu and Kashmir | 2961 | 2754 | 207 |
| 4/10/2020 | Karnataka | 7975 | 7673 | 207 |
| 4/10/2020 | Kerala | 13339 | 12335 | 364 |
| 4/10/2020 | Madhya Pradesh | 7049 | 4840 | 451 |
| 4/10/2020 | Maharashtra | 30000 | 28865 | 1135 |
| 4/10/2020 | Odisha | 3249 | 3201 | 48 |
| 4/10/2020 | Punjab | 3461 | 2972 | 151 |
| 4/10/2020 | Rajasthan | 22324 | 20673 | 520 |
| 4/10/2020 | Tamil Nadu | 8410 | 6838 | 911 |
| 4/11/2020 | Maharashtra | 31841 | 30477 | 1364 |

```
In [9]:  dim(data) #Dimension of the dataset variable
```

201 · 5

In [10]: `str(data) #Structure of the dataset`

```
'data.frame':   201 obs. of  5 variables:
 $ Date        : Factor w/ 55 levels "2/17/2020","2/18/2020",..: 1 2 3 4 5 6
7 8 9 10 ...
 $ State       : Factor w/ 13 levels "Andhra Pradesh",..: 7 7 7 7 7 7 7 7 7 7
...
 $ TotalSamples: int  423 425 432 433 437 441 444 449 451 460 ...
 $ Negative    : int  406 420 423 423 426 436 436 441 446 451 ...
 $ Positive    : int  3 3 3 3 3 3 3 3 3 3 ...
```

In [15]: `summary(data)`

```
      Date                State      TotalSamples       Negative
 4/10/2020: 13   Kerala        :54   Min.   :   49   Min.   :   49.0
 4/8/2020 : 12   Tamil Nadu    :41   1st Qu.:  731   1st Qu.:  608.2
 4/9/2020 : 12   Karnataka     :29   Median : 2297   Median : 1851.0
 4/5/2020 : 11   Delhi         :14   Mean   : 3961   Mean   : 3518.7
 4/6/2020 : 10   Madhya Pradesh:10   3rd Qu.: 5015   3rd Qu.: 4071.5
 4/7/2020 : 10   Odisha        :10   Max.   :31841   Max.   :30477.0
 (Other)  :133   (Other)       :43                   NA's   :3
    Positive
 Min.   :   0.0
 1st Qu.:   9.0
 Median :  72.0
 Mean   : 171.8
 3rd Qu.: 234.0
 Max.   :1364.0
```

In [17]: `sum(data$TotalSamples) #Total number of COVID testing samples`

796152

In [19]: `sum(data$Positive) #Total number of Positive cases`

34533

In [20]: `sum(data$Negative)`

<NA>

In [14]: `any(is.na(data)) #any missing value present in my dataset`

TRUE

In [22]: `data_1 = na.omit(data) #Removing all the missing value from the dataset`

In [25]: `sum(data_1$Negative) #Total number of negative cases`

696703

In [28]: `any(is.na(data_1)) #any missing value present in my dataset`

FALSE

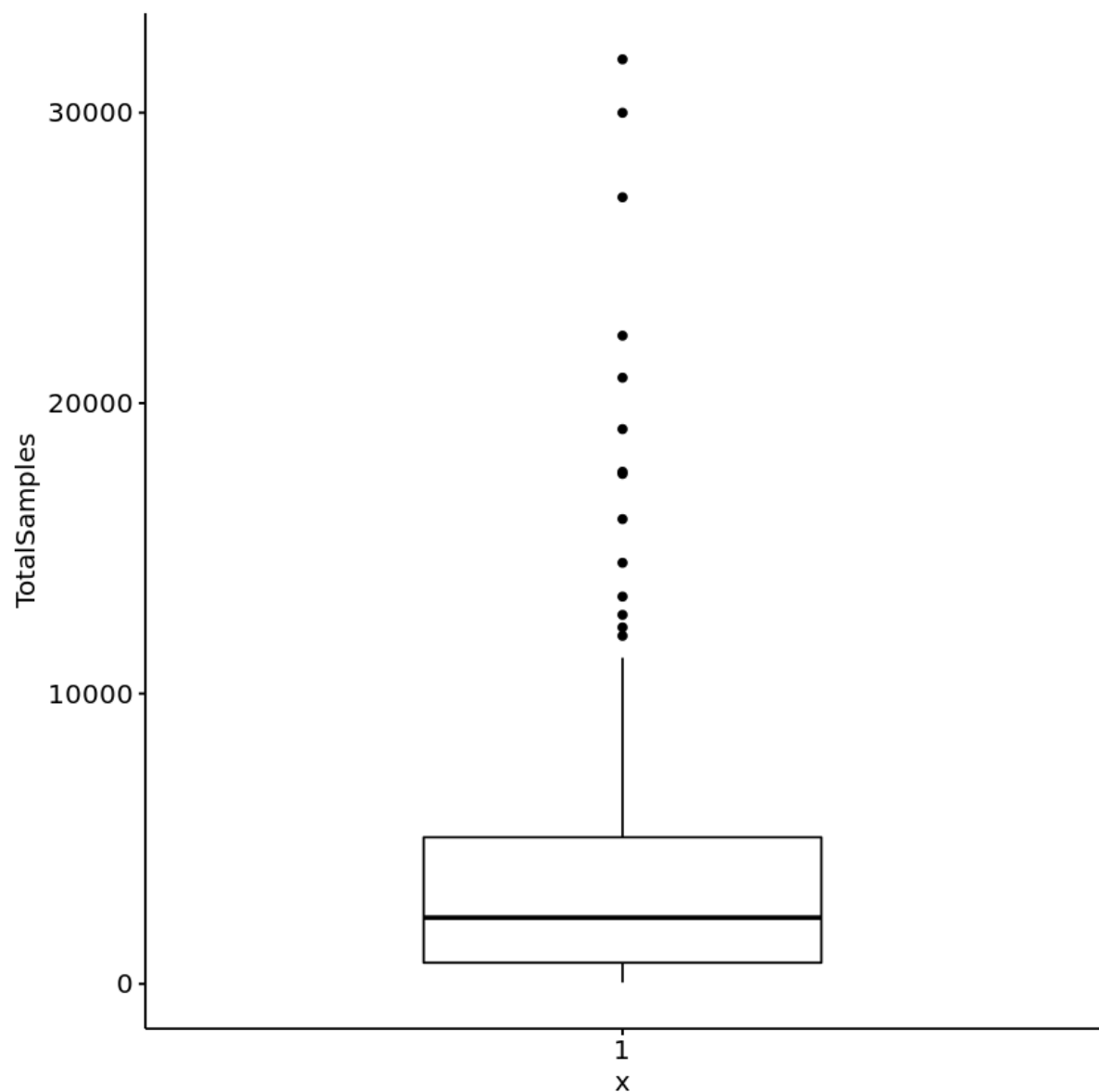In [31]: `install.packages("ggplot2")`

```
Installing package into '/srv/rlibs'
(as 'lib' is unspecified)
```
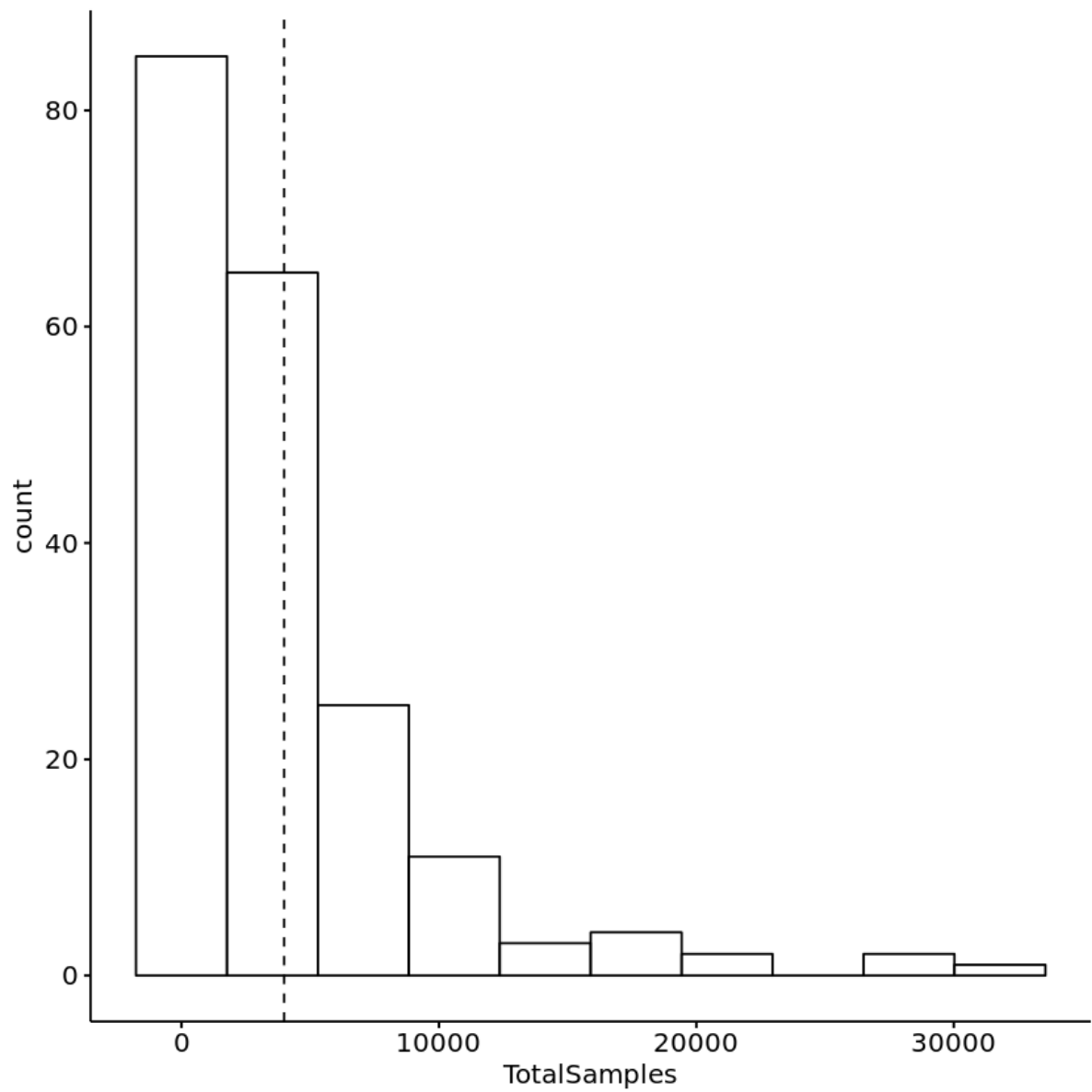
In [34]: `install.packages("ggpubr")`

```
Installing package into '/srv/rlibs'
(as 'lib' is unspecified)
```
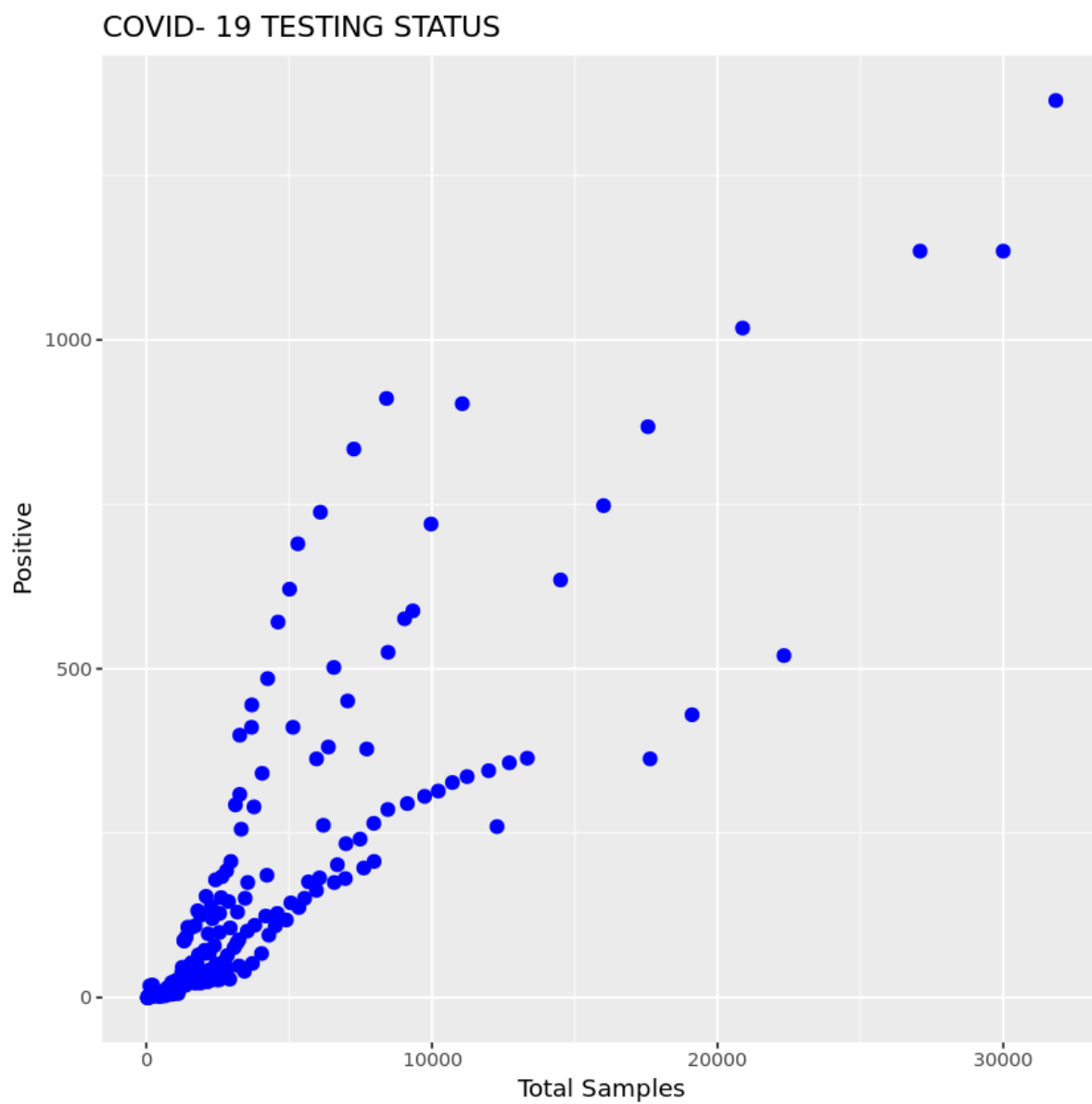
In [36]: `library(ggplot2)`
`library(ggpubr)`

In [39]: `ggboxplot(data_1, y = "TotalSamples", width = 0.5)`

In [42]: `gghistogram(data_1, x = "TotalSamples", bins = 10,add = "mean")`

In [45]: 
```
ggplot(data_1, aes(x= TotalSamples, y = Positive)) + geom_point(size = 2.5, co
lor="blue") + xlab("Total Samples") +
ylab("Positive") + ggtitle("COVID- 19 TESTING STATUS")
```



COVID- 19 TESTING STATUS

In [47]:
```r
ggplot(data_1, aes(State, Positive)) + geom_bar( stat = "identity") +
theme(axis.text.x = element_text(angle = 70, vjust = 0.5, color = "blue")) +
xlab("State") + ylab("Positive")+ggtitle("Statewise_Positive Tests")
```

### Statewise_Positive Tests
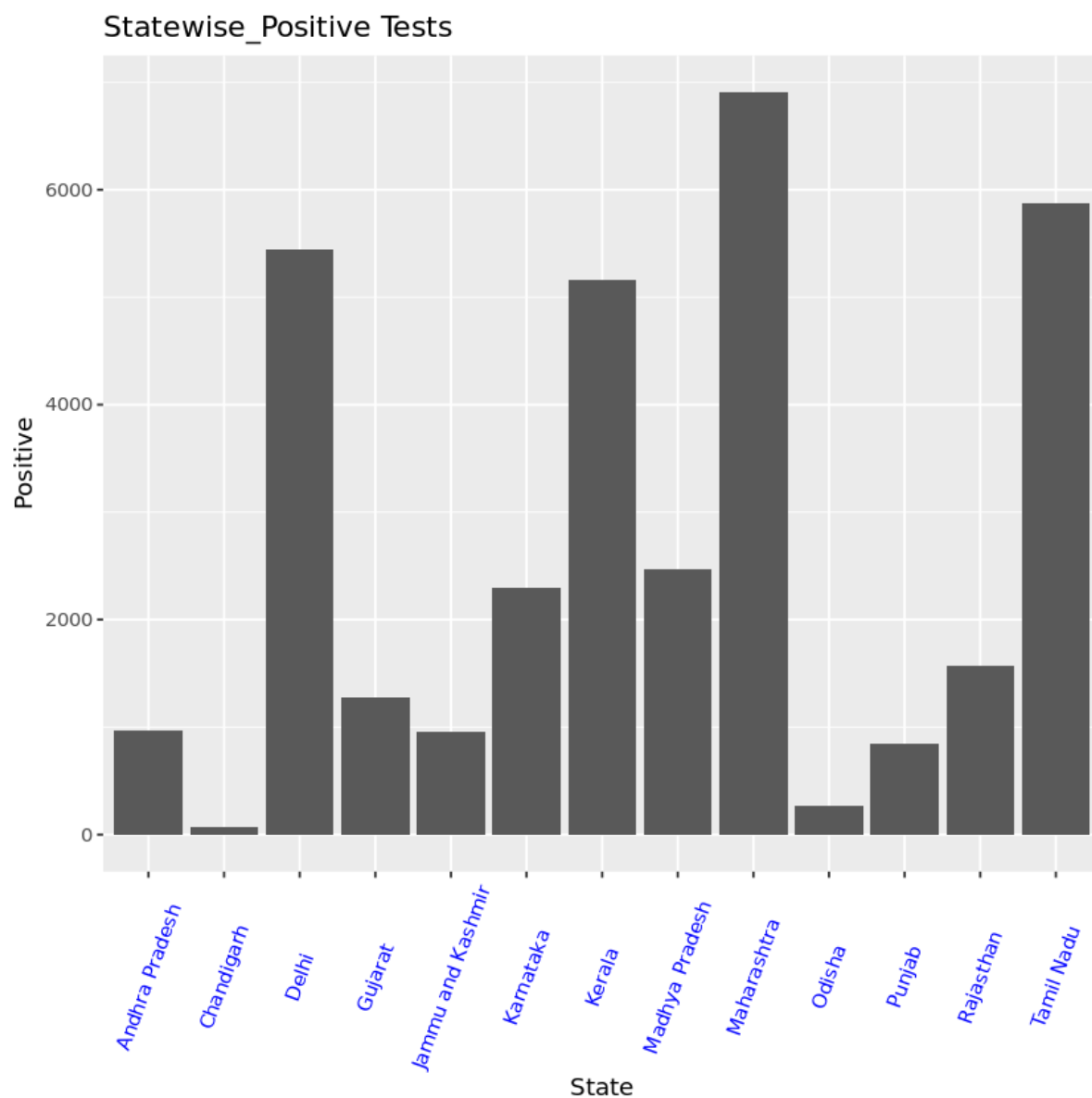
In [49]:
```
ggplot(data_1, aes(Date, TotalSamples)) + geom_bar( stat = "identity") +
theme(axis.text.x = element_text(angle = 70, vjust = 0.5, color = "red")) +
xlab("Date") + ylab("TotalSamples")+ggtitle("Date_wise_Total Samples Cases")
```
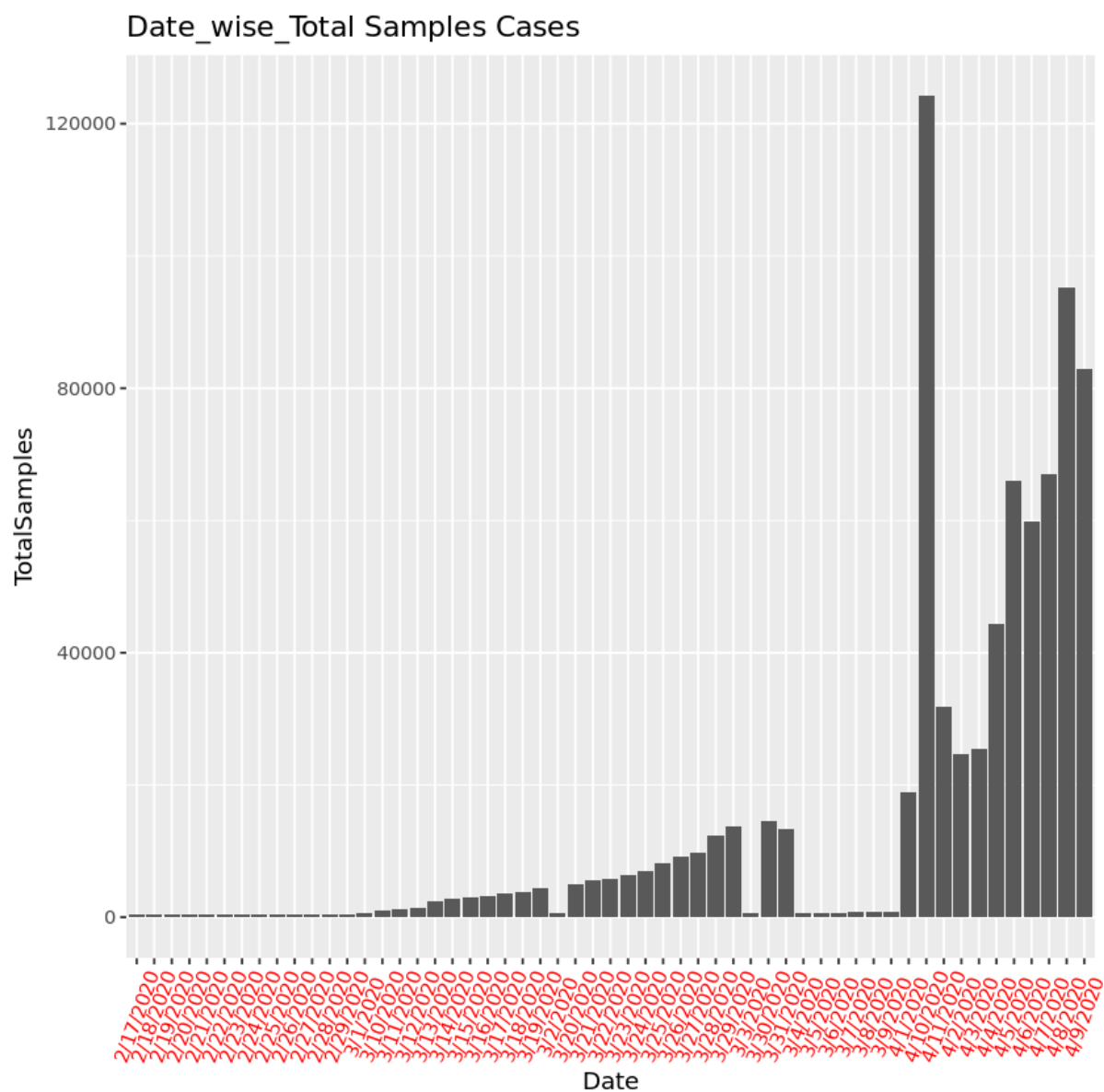
```
In [50]:  ggplot(data_1, aes(x = Date, y = State)) +   geom_point(aes(color = Date)) +
          geom_smooth(method = 'lm')
```
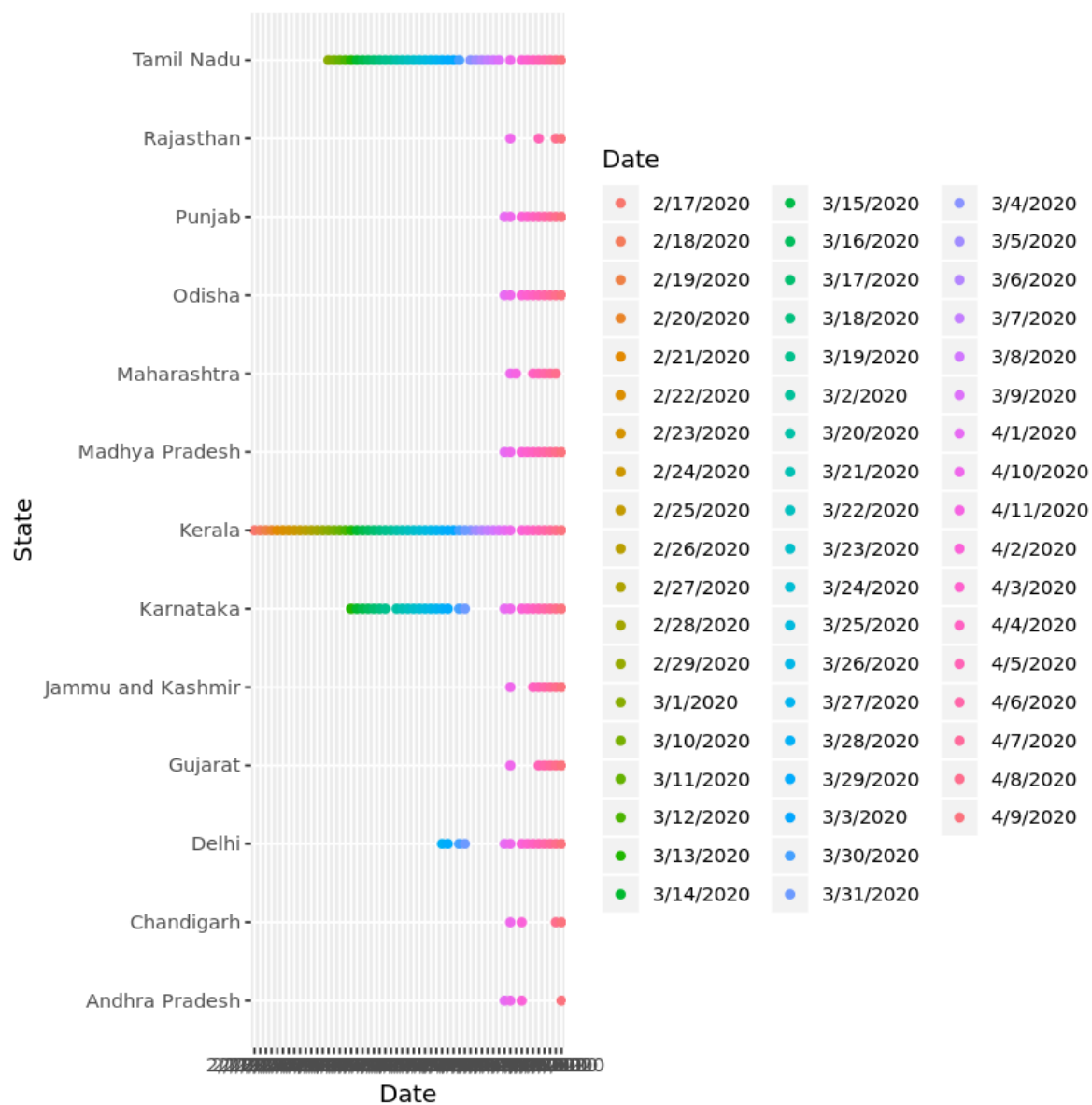
In [52]:
```r
layer_point <- geom_point(
  mapping = aes(x = State, y = Date, color = Date),
  data = data_1,
  size = 3
)
ggplot() + layer_point
```