Forest_cover_Data

Durga Gaddam

November 21, 2016

Objective

The main objective of this project is to visualize the data of forest cover type provided by kaggle and predict the cover type from the train data set provided.

This is a classification problem, We can use any of the following classification algorithms

1) Decision Trees
2) Random Forests
3) Support Vector Machines
4) Neural Networks
5) K Nearset Neighbour
6) Naive Bayes
7) XGBoost

library(data.table) library(dplyr) install.packages("moments") library(moments) install.packages("PerformanceAnalytics") library(PerformanceAnalytics) library(ggthemes) library(ggplot2) library(RColorBrewer) library(vioplot) install.packages("vioplot") install.packages("xgboost") setwd("F:/R PRACTICE/ForestCover")

```
require(data.table)
```

```
## Loading required package: data.table
```

```
## Warning: package 'data.table' was built under R version 3.3.1
```

```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.3.1
```

```
require(moments)
```

```
## Loading required package: moments
```

```
## Warning: package 'moments' was built under R version 3.3.2
```

```
require(xgboost)
```

```
## Loading required package: xgboost
```

```
## Warning: package 'xgboost' was built under R version 3.3.2
```

```
train <- fread("train.csv", stringsAsFactors = FALSE)
train <- train[, Id:=NULL, ]
train <- train[, Soil_Type15:=NULL]
train <- train[, Soil_Type7:=NULL]
is.data.table(train)
```

```
## [1] TRUE
```

```
names(train)
```

```
##  [1] "Elevation"
##  [2] "Aspect"
##  [3] "Slope"
##  [4] "Horizontal_Distance_To_Hydrology"
##  [5] "Vertical_Distance_To_Hydrology"
##  [6] "Horizontal_Distance_To_Roadways"
##  [7] "Hillshade_9am"
##  [8] "Hillshade_Noon"
##  [9] "Hillshade_3pm"
## [10] "Horizontal_Distance_To_Fire_Points"
## [11] "Wilderness_Area1"
## [12] "Wilderness_Area2"
## [13] "Wilderness_Area3"
## [14] "Wilderness_Area4"
## [15] "Soil_Type1"
## [16] "Soil_Type2"
## [17] "Soil_Type3"
## [18] "Soil_Type4"
## [19] "Soil_Type5"
## [20] "Soil_Type6"
## [21] "Soil_Type8"
## [22] "Soil_Type9"
## [23] "Soil_Type10"
## [24] "Soil_Type11"
## [25] "Soil_Type12"
## [26] "Soil_Type13"
## [27] "Soil_Type14"
## [28] "Soil_Type16"
## [29] "Soil_Type17"
## [30] "Soil_Type18"
## [31] "Soil_Type19"
## [32] "Soil_Type20"
## [33] "Soil_Type21"
## [34] "Soil_Type22"
## [35] "Soil_Type23"
## [36] "Soil_Type24"
## [37] "Soil_Type25"
## [38] "Soil_Type26"
## [39] "Soil_Type27"
## [40] "Soil_Type28"
## [41] "Soil_Type29"
## [42] "Soil_Type30"
## [43] "Soil_Type31"
```

```
## [44] "Soil_Type32"
## [45] "Soil_Type33"
## [46] "Soil_Type34"
## [47] "Soil_Type35"
## [48] "Soil_Type36"
## [49] "Soil_Type37"
## [50] "Soil_Type38"
## [51] "Soil_Type39"
## [52] "Soil_Type40"
## [53] "Cover_Type"
```

Here the Cover_Type is the outcome variable that we are interested in.

```
dim(train)
```

```
## [1] 15120    53
```

```
str(train)
```

```
## Classes 'data.table' and 'data.frame':   15120 obs. of  53 variables:
##  $ Elevation                         : int  2596 2590 2804 2785 2595 2579 2606 2605 2617 2612 ...
##  $ Aspect                            : int  51 56 139 155 45 132 45 49 45 59 ...
##  $ Slope                             : int  3 2 9 18 2 6 7 4 9 10 ...
##  $ Horizontal_Distance_To_Hydrology  : int  258 212 268 242 153 300 270 234 240 247 ...
##  $ Vertical_Distance_To_Hydrology    : int  0 -6 65 118 -1 -15 5 7 56 11 ...
##  $ Horizontal_Distance_To_Roadways   : int  510 390 3180 3090 391 67 633 573 666 636 ...
##  $ Hillshade_9am                     : int  221 220 234 238 220 230 222 222 223 228 ...
##  $ Hillshade_Noon                    : int  232 235 238 238 234 237 225 230 221 219 ...
##  $ Hillshade_3pm                     : int  148 151 135 122 150 140 138 144 133 124 ...
##  $ Horizontal_Distance_To_Fire_Points: int  6279 6225 6121 6211 6172 6031 6256 6228 6244 6230 ...
##  $ Wilderness_Area1                  : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Wilderness_Area2                  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Wilderness_Area3                  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Wilderness_Area4                  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type1                        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type2                        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type3                        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type4                        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type5                        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type6                        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type8                        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type9                        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type10                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type11                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type12                       : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ Soil_Type13                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type14                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type16                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type17                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type18                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type19                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type20                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type21                       : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ Soil_Type22                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type23                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type24                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type25                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type26                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type27                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type28                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type29                       : int  1 1 0 0 1 1 1 1 1 1 ...
##  $ Soil_Type30                       : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ Soil_Type31                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type32                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type33                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type34                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type35                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type36                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type37                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type38                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type39                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Soil_Type40                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Cover_Type                        : int  5 5 2 2 5 2 5 5 5 5 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```
summary(train)
```

```
##    Elevation        Aspect          Slope
##  Min.   :1863   Min.   :  0.0   Min.   : 0.0
##  1st Qu.:2376   1st Qu.: 65.0   1st Qu.:10.0
##  Median :2752   Median :126.0   Median :15.0
##  Mean   :2749   Mean   :156.7   Mean   :16.5
##  3rd Qu.:3104   3rd Qu.:261.0   3rd Qu.:22.0
##  Max.   :3849   Max.   :360.0   Max.   :52.0
##  Horizontal_Distance_To_Hydrology Vertical_Distance_To_Hydrology
##  Min.   :   0.0                   Min.   :-146.00
##  1st Qu.:  67.0                   1st Qu.:   5.00
##  Median : 180.0                   Median :  32.00
##  Mean   : 227.2                   Mean   :  51.08
##  3rd Qu.: 330.0                   3rd Qu.:  79.00
##  Max.   :1343.0                   Max.   : 554.00
##  Horizontal_Distance_To_Roadways Hillshade_9am   Hillshade_Noon
##  Min.   :   0                    Min.   :  0.0   Min.   : 99
##  1st Qu.: 764                    1st Qu.:196.0   1st Qu.:207
##  Median :1316                    Median :220.0   Median :223
##  Mean   :1714                    Mean   :212.7   Mean   :219
##  3rd Qu.:2270                    3rd Qu.:235.0   3rd Qu.:235
##  Max.   :6890                    Max.   :254.0   Max.   :254
##  Hillshade_3pm   Horizontal_Distance_To_Fire_Points Wilderness_Area1
##  Min.   :  0.0   Min.   :   0                       Min.   :0.0000
##  1st Qu.:106.0   1st Qu.: 730                      1st Qu.:0.0000
##  Median :138.0   Median :1256                      Median :0.0000
##  Mean   :135.1   Mean   :1511                      Mean   :0.2379
##  3rd Qu.:167.0   3rd Qu.:1988                      3rd Qu.:0.0000
##  Max.   :248.0   Max.   :6993                      Max.   :1.0000
##  Wilderness_Area2 Wilderness_Area3 Wilderness_Area4   Soil_Type1
##  Min.   :0.000    Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
```

```
##  1st Qu.:0.000    1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.000    Median :0.0000   Median :0.0000   Median :0.00000
##  Mean   :0.033    Mean   :0.4199   Mean   :0.3092   Mean   :0.02348
##  3rd Qu.:0.000    3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.00000
##  Max.   :1.000    Max.   :1.0000   Max.   :1.0000   Max.   :1.00000
##    Soil_Type2       Soil_Type3        Soil_Type4        Soil_Type5
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0412   Mean   :0.06362   Mean   :0.05575   Mean   :0.01091
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##    Soil_Type6        Soil_Type8         Soil_Type9         Soil_Type10
##  Min.   :0.00000   Min.   :0.00e+00   Min.   :0.0000000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00e+00   1st Qu.:0.0000000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00e+00   Median :0.0000000   Median :0.0000
##  Mean   :0.04299   Mean   :6.61e-05   Mean   :0.0006614   Mean   :0.1417
##  3rd Qu.:0.00000   3rd Qu.:0.00e+00   3rd Qu.:0.0000000   3rd Qu.:0.0000
##  Max.   :1.00000   Max.   :1.00e+00   Max.   :1.0000000   Max.   :1.0000
##   Soil_Type11       Soil_Type12       Soil_Type13       Soil_Type14
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.02685   Mean   :0.01501   Mean   :0.03148   Mean   :0.01118
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##   Soil_Type16       Soil_Type17       Soil_Type18        Soil_Type19
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.000000   Min.   :0.000000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.000000
##  Median :0.00000   Median :0.00000   Median :0.000000   Median :0.000000
##  Mean   :0.00754   Mean   :0.04048   Mean   :0.003968   Mean   :0.003042
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.000000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.000000   Max.   :1.000000
##   Soil_Type20        Soil_Type21        Soil_Type22       Soil_Type23
##  Min.   :0.000000   Min.   :0.000000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.000000   Median :0.000000   Median :0.00000   Median :0.00000
##  Mean   :0.009193   Mean   :0.001058   Mean   :0.02282   Mean   :0.05007
##  3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.000000   Max.   :1.000000   Max.   :1.00000   Max.   :1.00000
##   Soil_Type24     Soil_Type25        Soil_Type26        Soil_Type27
##  Min.   :0.000   Min.   :0.00e+00   Min.   :0.000000   Min.   :0.0000000
##  1st Qu.:0.000   1st Qu.:0.00e+00   1st Qu.:0.000000   1st Qu.:0.0000000
##  Median :0.000   Median :0.00e+00   Median :0.000000   Median :0.0000000
##  Mean   :0.017   Mean   :6.61e-05   Mean   :0.003571   Mean   :0.0009921
##  3rd Qu.:0.000   3rd Qu.:0.00e+00   3rd Qu.:0.000000   3rd Qu.:0.0000000
##  Max.   :1.000   Max.   :1.00e+00   Max.   :1.000000   Max.   :1.0000000
##   Soil_Type28         Soil_Type29       Soil_Type30       Soil_Type31
##  Min.   :0.0000000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0005952   Mean   :0.08538   Mean   :0.04795   Mean   :0.02196
##  3rd Qu.:0.0000000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
```

```
##   Soil_Type32       Soil_Type33       Soil_Type34        Soil_Type35
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.000000   Min.   :0.000000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.000000
##  Median :0.00000   Median :0.00000   Median :0.000000   Median :0.000000
##  Mean   :0.04563   Mean   :0.04074   Mean   :0.001455   Mean   :0.006746
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.000000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.000000   Max.   :1.000000
##   Soil_Type36         Soil_Type37        Soil_Type38
##  Min.   :0.0000000   Min.   :0.000000   Min.   :0.00000
##  1st Qu.:0.0000000   1st Qu.:0.000000   1st Qu.:0.00000
##  Median :0.0000000   Median :0.000000   Median :0.00000
##  Mean   :0.0006614   Mean   :0.002249   Mean   :0.04815
##  3rd Qu.:0.0000000   3rd Qu.:0.000000   3rd Qu.:0.00000
##  Max.   :1.0000000   Max.   :1.000000   Max.   :1.00000
##   Soil_Type39       Soil_Type40        Cover_Type
##  Min.   :0.00000   Min.   :0.00000   Min.   :1
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:2
##  Median :0.00000   Median :0.00000   Median :4
##  Mean   :0.04345   Mean   :0.03036   Mean   :4
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:6
##  Max.   :1.00000   Max.   :1.00000   Max.   :7
```

```
skewness(train)
```

```
##                            Elevation                               Aspect
##                            0.0756322                            0.4508906
##                                Slope       Horizontal_Distance_To_Hydrology
##                            0.5236064                            1.4879049
##      Vertical_Distance_To_Hydrology        Horizontal_Distance_To_Roadways
##                            1.5376231                            1.2476869
##                         Hillshade_9am                        Hillshade_Noon
##                           -1.0935721                           -0.9531371
##                        Hillshade_3pm     Horizontal_Distance_To_Fire_Points
##                           -0.3407934                            1.6169384
##                    Wilderness_Area1                     Wilderness_Area2
##                            1.2311219                            5.2282626
##                    Wilderness_Area3                     Wilderness_Area4
##                            0.3245614                            0.8257162
##                         Soil_Type1                           Soil_Type2
##                            6.2940919                            4.6165612
##                         Soil_Type3                           Soil_Type4
##                            3.5756402                            3.8723364
##                         Soil_Type5                           Soil_Type6
##                            9.4152747                            4.5062689
##                         Soil_Type8                           Soil_Type9
##                          122.9512101                           38.8458577
##                        Soil_Type10                          Soil_Type11
##                            2.0552061                            5.8539706
##                        Soil_Type12                          Soil_Type13
##                            7.9764135                            5.3663033
##                        Soil_Type14                          Soil_Type16
##                            9.2993956                           11.3859201
##                        Soil_Type17                          Soil_Type18
##                            4.6634822                           15.7798601
```

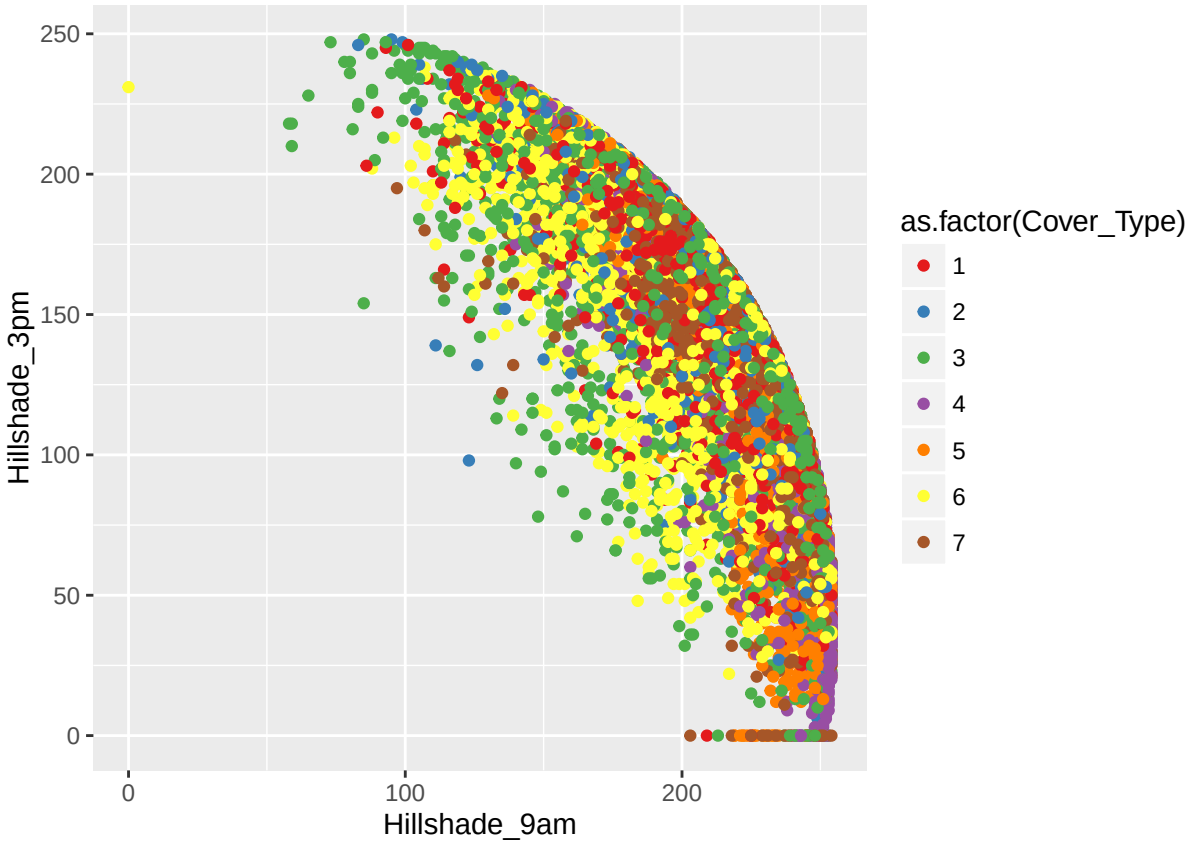
```
##                       Soil_Type19                       Soil_Type20
##                        18.0471245                        10.2852446
##                       Soil_Type21                       Soil_Type22
##                        30.6920358                         6.3913564
##                       Soil_Type23                       Soil_Type24
##                         4.1262919                         7.4732842
##                       Soil_Type25                       Soil_Type26
##                       122.9512101                        16.6434247
##                       Soil_Type27                       Soil_Type28
##                        31.7017506                        40.9511977
##                       Soil_Type29                       Soil_Type30
##                         2.9673567                         4.2314935
##                       Soil_Type31                       Soil_Type32
##                         6.5241566                         4.3544070
##                       Soil_Type33                       Soil_Type34
##                         4.6462808                        26.1586350
##                       Soil_Type35                       Soil_Type36
##                        12.0516425                        38.8458577
##                       Soil_Type37                       Soil_Type38
##                        21.0168536                         4.2213522
##                       Soil_Type39                       Soil_Type40
##                         4.4787414                         5.4747131
##                       Cover_Type
##                         0.0000000
```

```
table(train$Cover_Type)
```

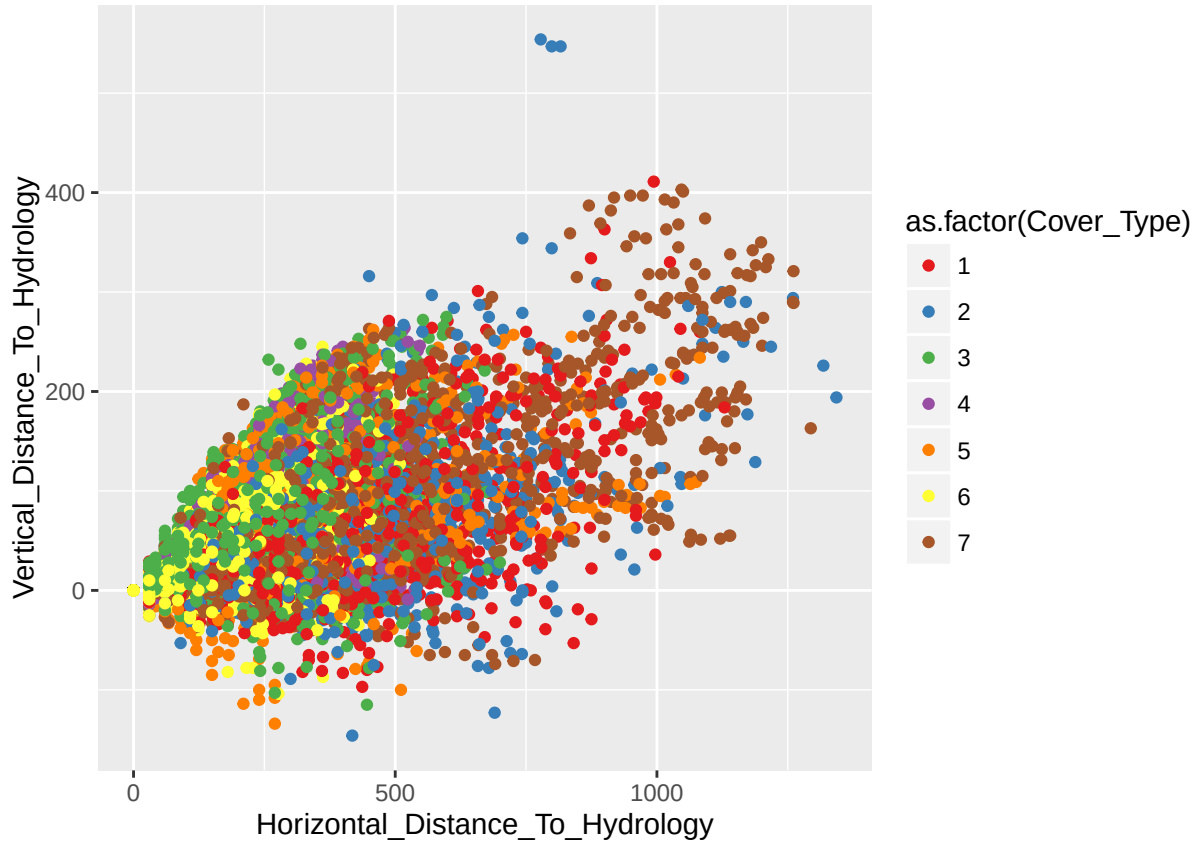
```
##
##    1    2    3    4    5    6    7
## 2160 2160 2160 2160 2160 2160 2160
```

Since Wilderness_Area and Solid_type are both binary we remove those features for correlation.

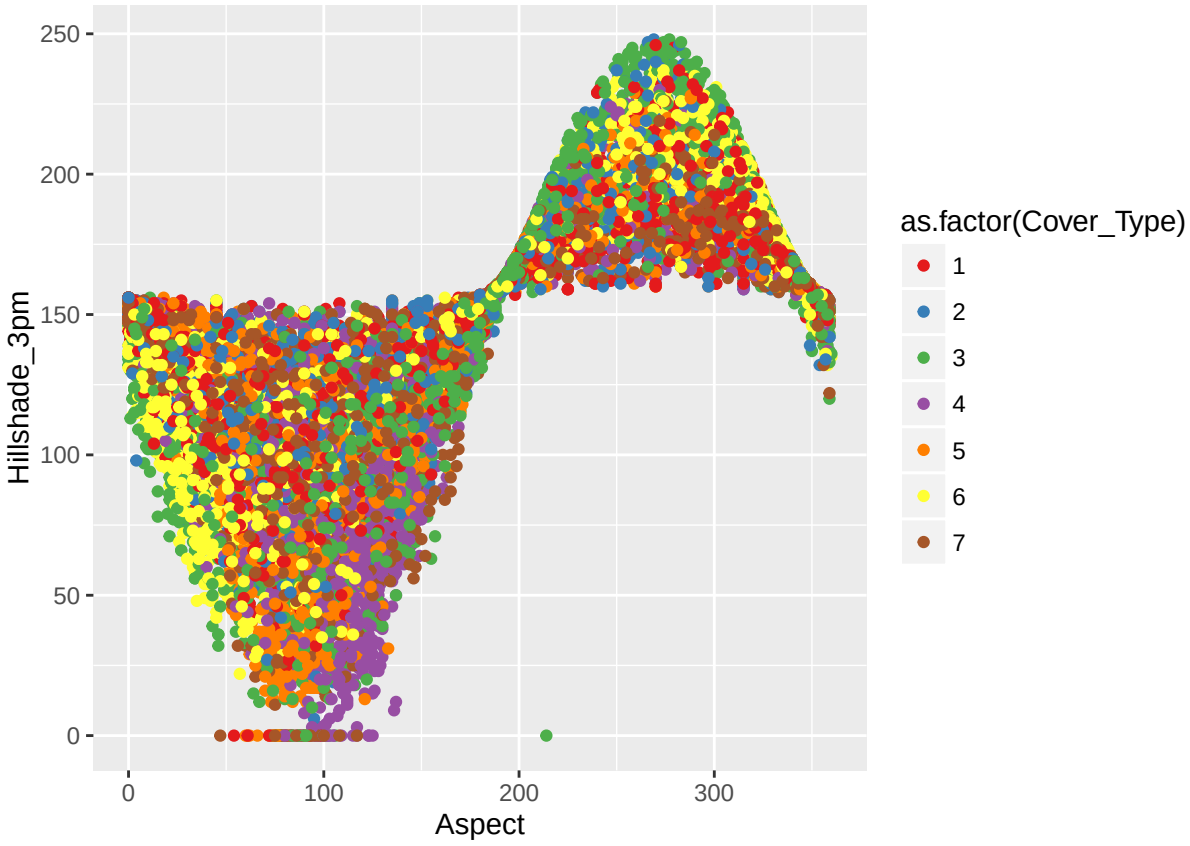
```
ggplot(train, aes(Hillshade_9am, Hillshade_3pm ))+
        geom_point(aes(colour= as.factor(Cover_Type)))+scale_color_brewer(palette="Set1")
```



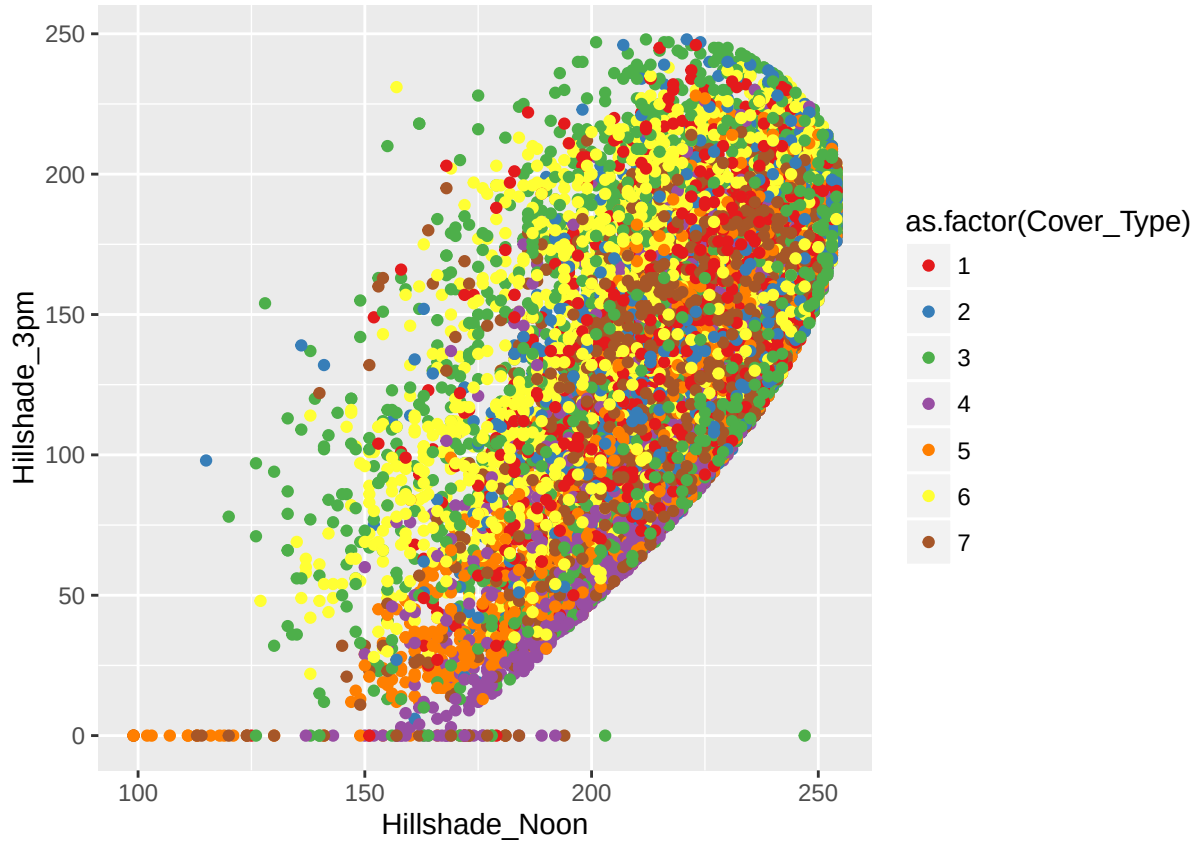

```
ggplot(train, aes(Horizontal_Distance_To_Hydrology,Vertical_Distance_To_Hydrology))+ geom_point(aes(col
```



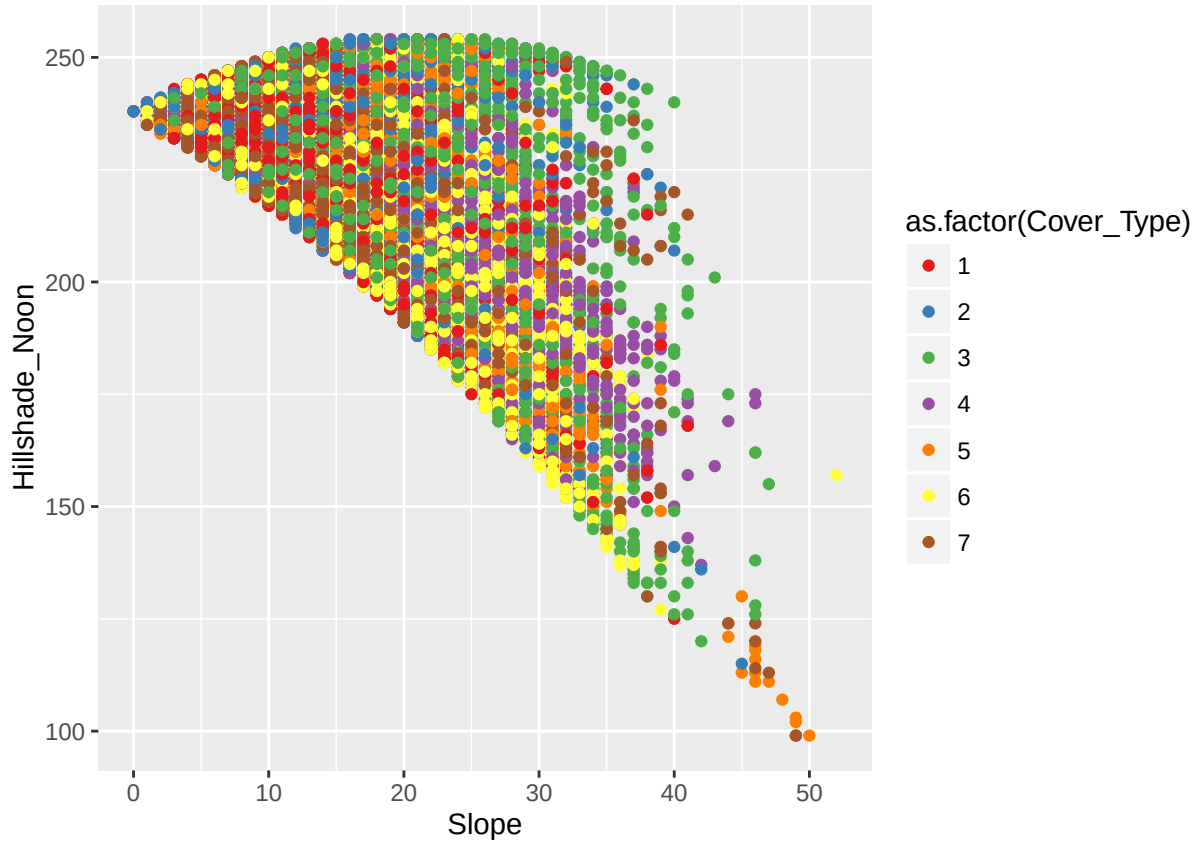

```
ggplot(train, aes(Aspect, Hillshade_3pm ))+
        geom_point(aes(colour= as.factor(Cover_Type)))+scale_color_brewer(palette="Set1")
```



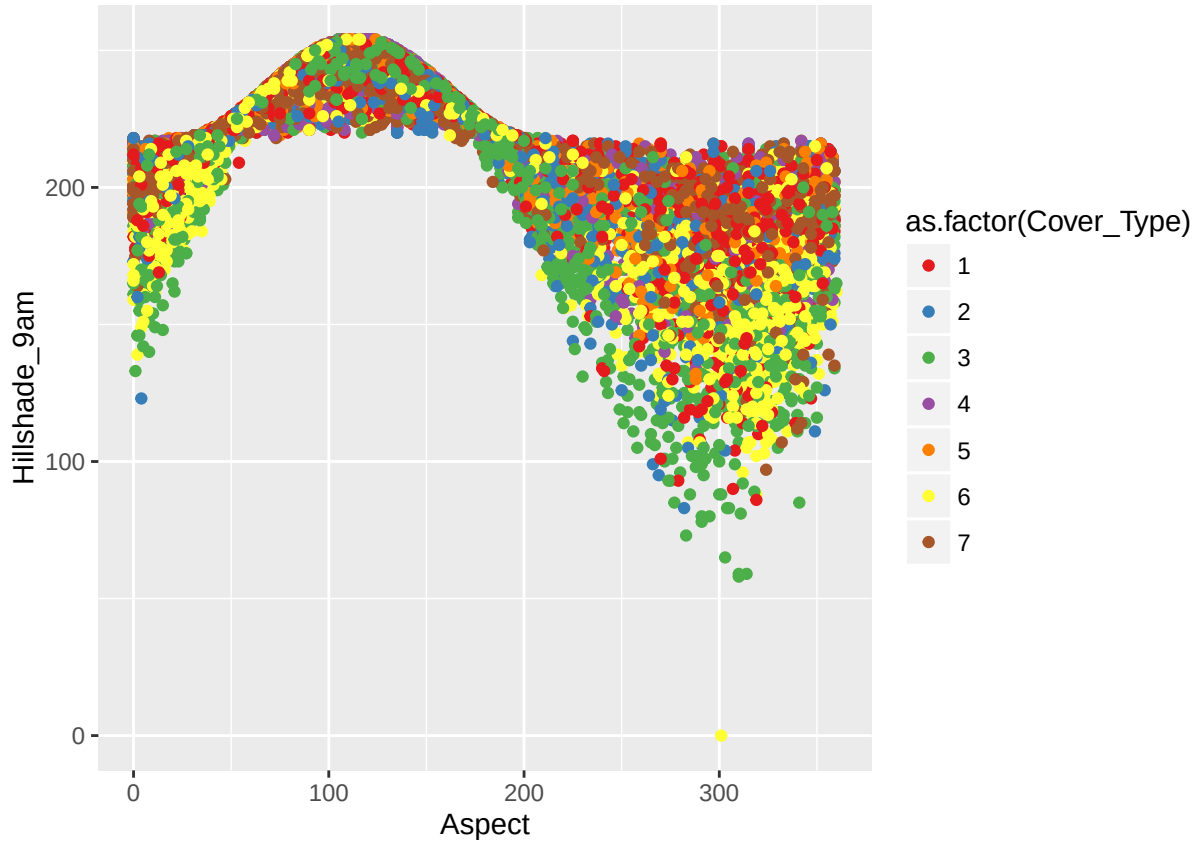

```
ggplot(train, aes(Hillshade_Noon, Hillshade_3pm ))+
        geom_point(aes(colour= as.factor(Cover_Type)))+scale_color_brewer(palette="Set1")
```



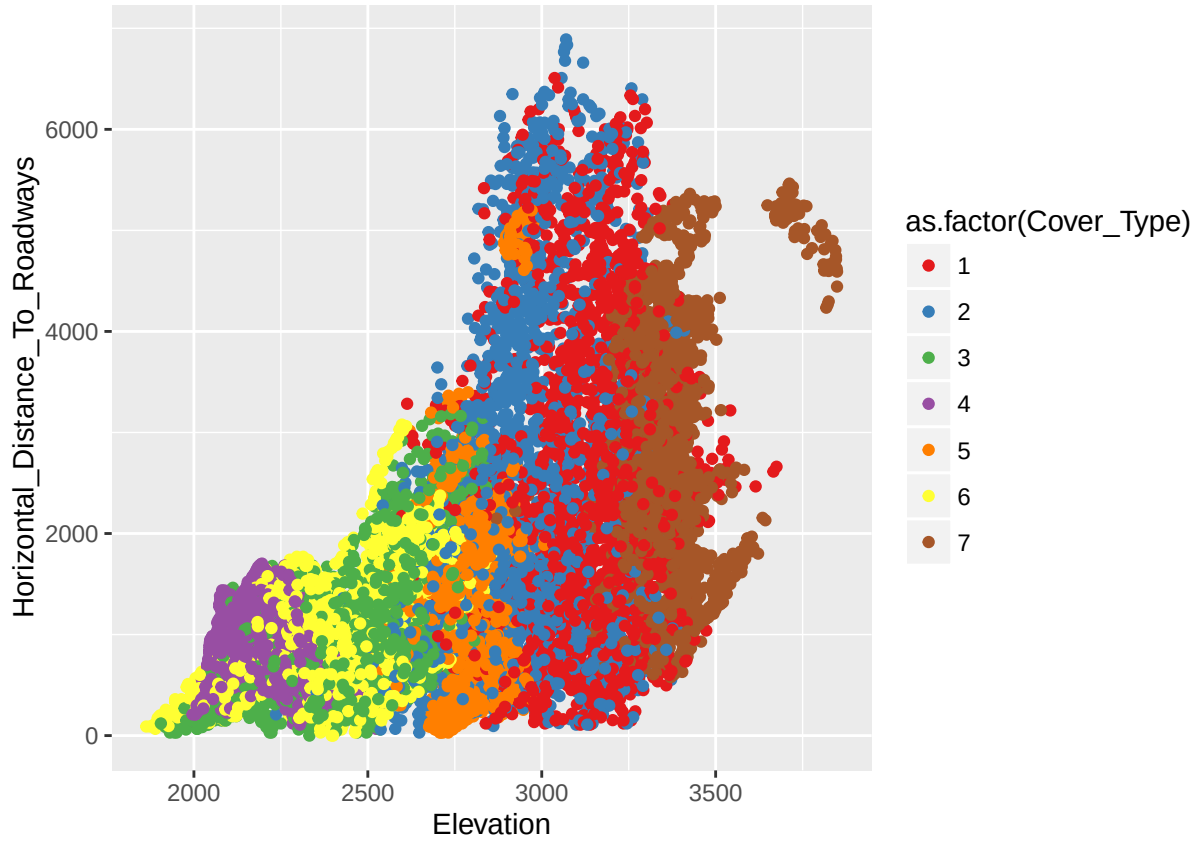

```
ggplot(train, aes(Slope, Hillshade_Noon))+
        geom_point(aes(colour= as.factor(Cover_Type)))+scale_color_brewer(palette="Set1")
```




```
ggplot(train, aes(Aspect, Hillshade_9am))+
       geom_point(aes(colour= as.factor(Cover_Type)))+scale_color_brewer(palette="Set1")
```




```
ggplot(train, aes(Elevation,Horizontal_Distance_To_Roadways))+
        geom_point(aes(colour= as.factor(Cover_Type)))+scale_color_brewer(palette="Set1")
```

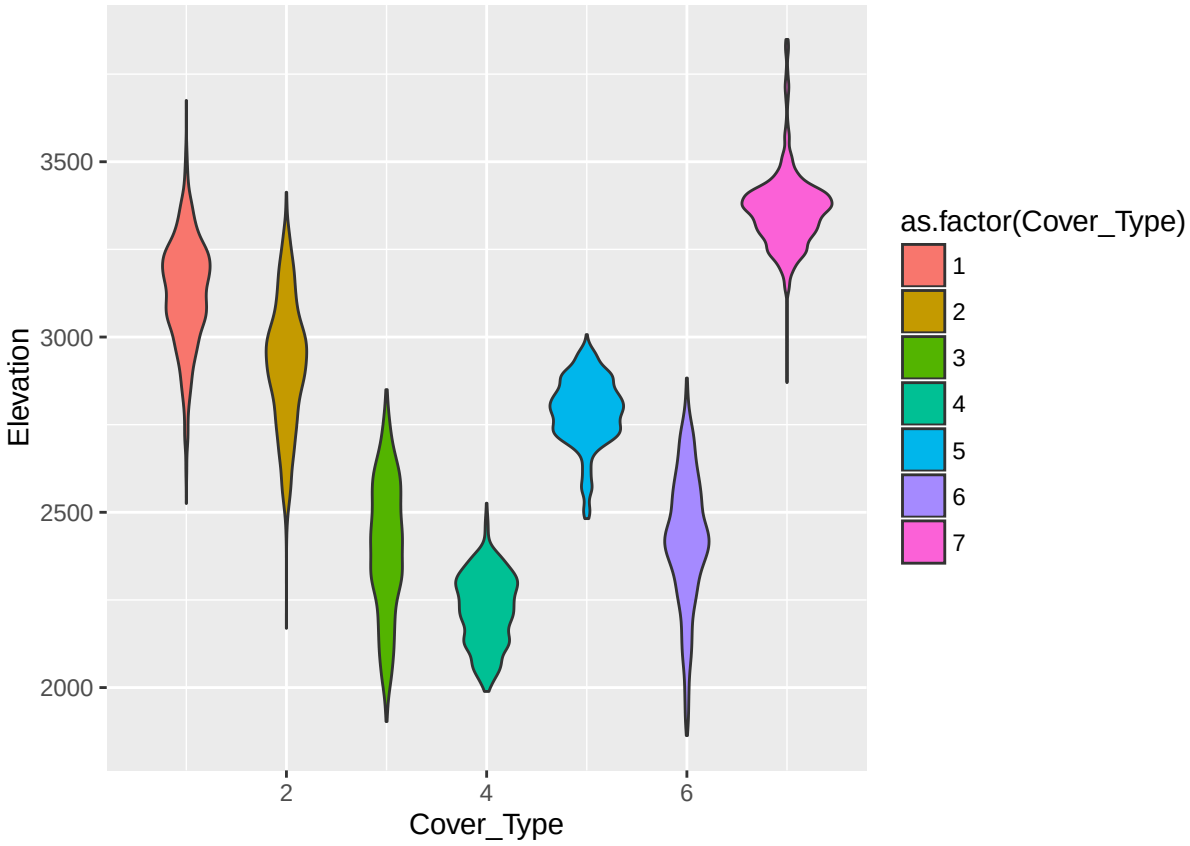


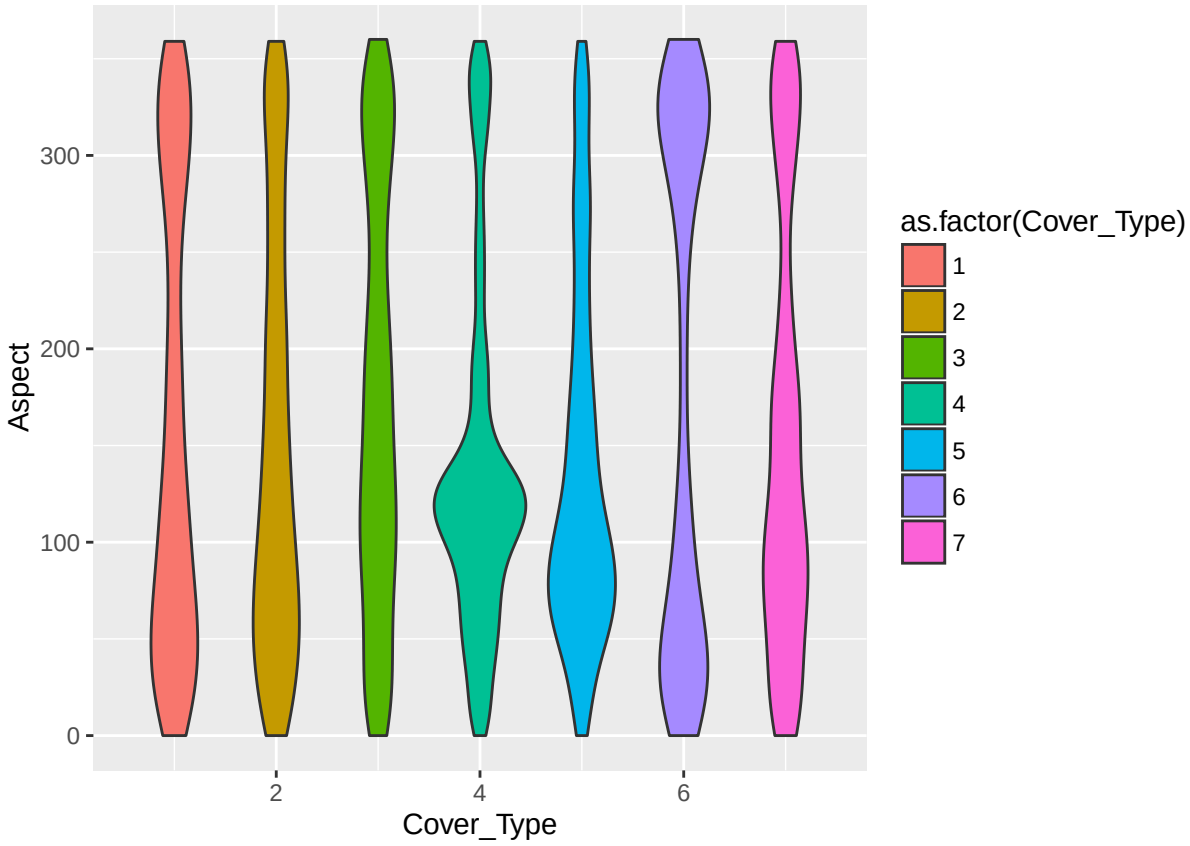

```
xm <- names(train)
xm
```

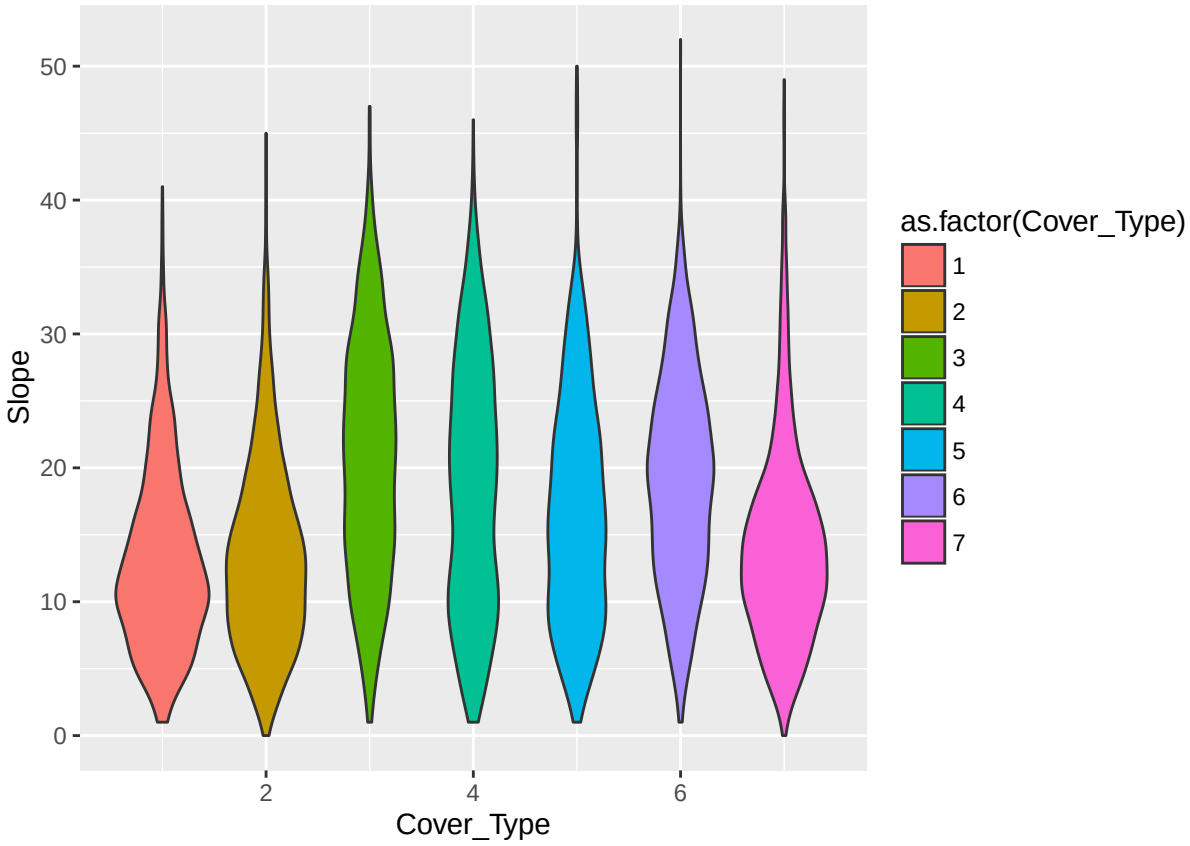
```
##  [1] "Elevation"
##  [2] "Aspect"
##  [3] "Slope"
##  [4] "Horizontal_Distance_To_Hydrology"
##  [5] "Vertical_Distance_To_Hydrology"
##  [6] "Horizontal_Distance_To_Roadways"
##  [7] "Hillshade_9am"
##  [8] "Hillshade_Noon"
##  [9] "Hillshade_3pm"
## [10] "Horizontal_Distance_To_Fire_Points"
## [11] "Wilderness_Area1"
## [12] "Wilderness_Area2"
## [13] "Wilderness_Area3"
## [14] "Wilderness_Area4"
## [15] "Soil_Type1"
## [16] "Soil_Type2"
## [17] "Soil_Type3"
## [18] "Soil_Type4"
## [19] "Soil_Type5"
## [20] "Soil_Type6"
## [21] "Soil_Type8"
## [22] "Soil_Type9"
## [23] "Soil_Type10"
```

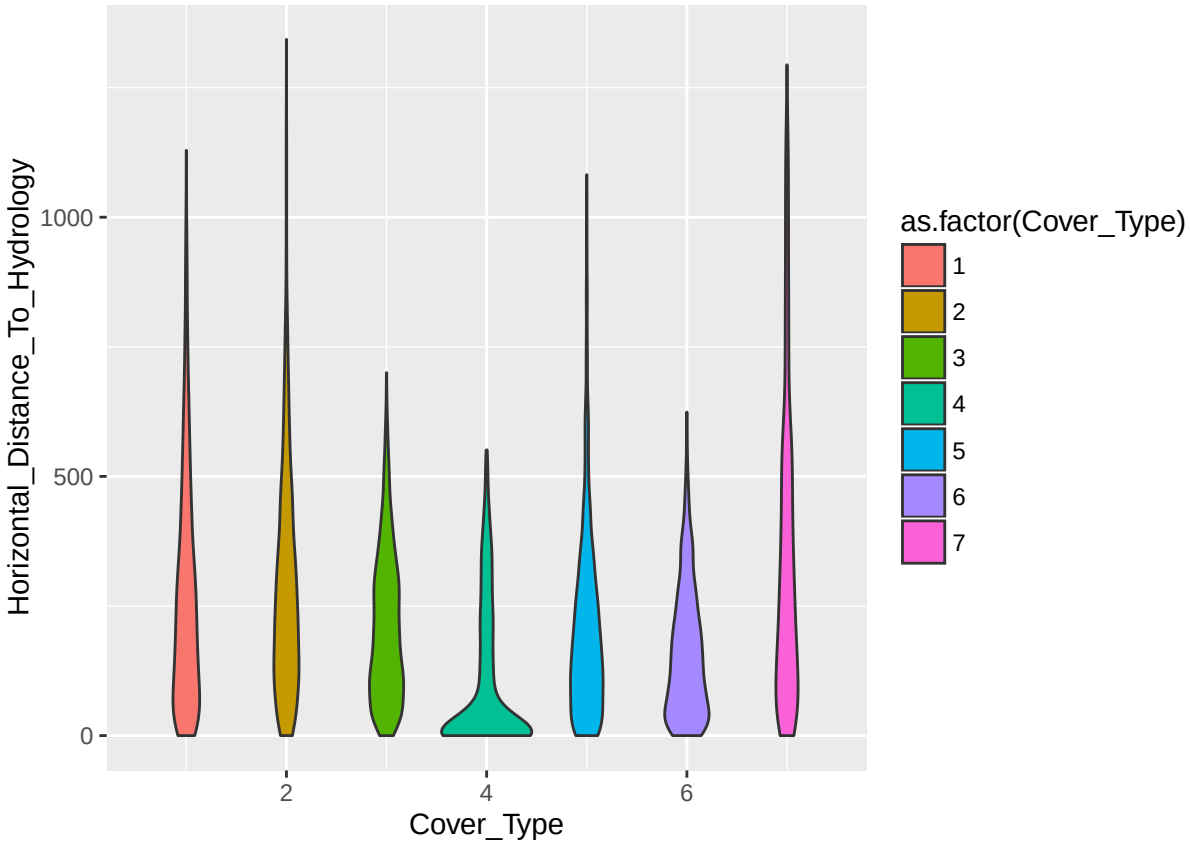
```
## [24] "Soil_Type11"
## [25] "Soil_Type12"
## [26] "Soil_Type13"
## [27] "Soil_Type14"
## [28] "Soil_Type16"
## [29] "Soil_Type17"
## [30] "Soil_Type18"
## [31] "Soil_Type19"
## [32] "Soil_Type20"
## [33] "Soil_Type21"
## [34] "Soil_Type22"
## [35] "Soil_Type23"
## [36] "Soil_Type24"
## [37] "Soil_Type25"
## [38] "Soil_Type26"
## [39] "Soil_Type27"
## [40] "Soil_Type28"
## [41] "Soil_Type29"
## [42] "Soil_Type30"
## [43] "Soil_Type31"
## [44] "Soil_Type32"
## [45] "Soil_Type33"
## [46] "Soil_Type34"
## [47] "Soil_Type35"
## [48] "Soil_Type36"
## [49] "Soil_Type37"
## [50] "Soil_Type38"
## [51] "Soil_Type39"
## [52] "Soil_Type40"
## [53] "Cover_Type"
```

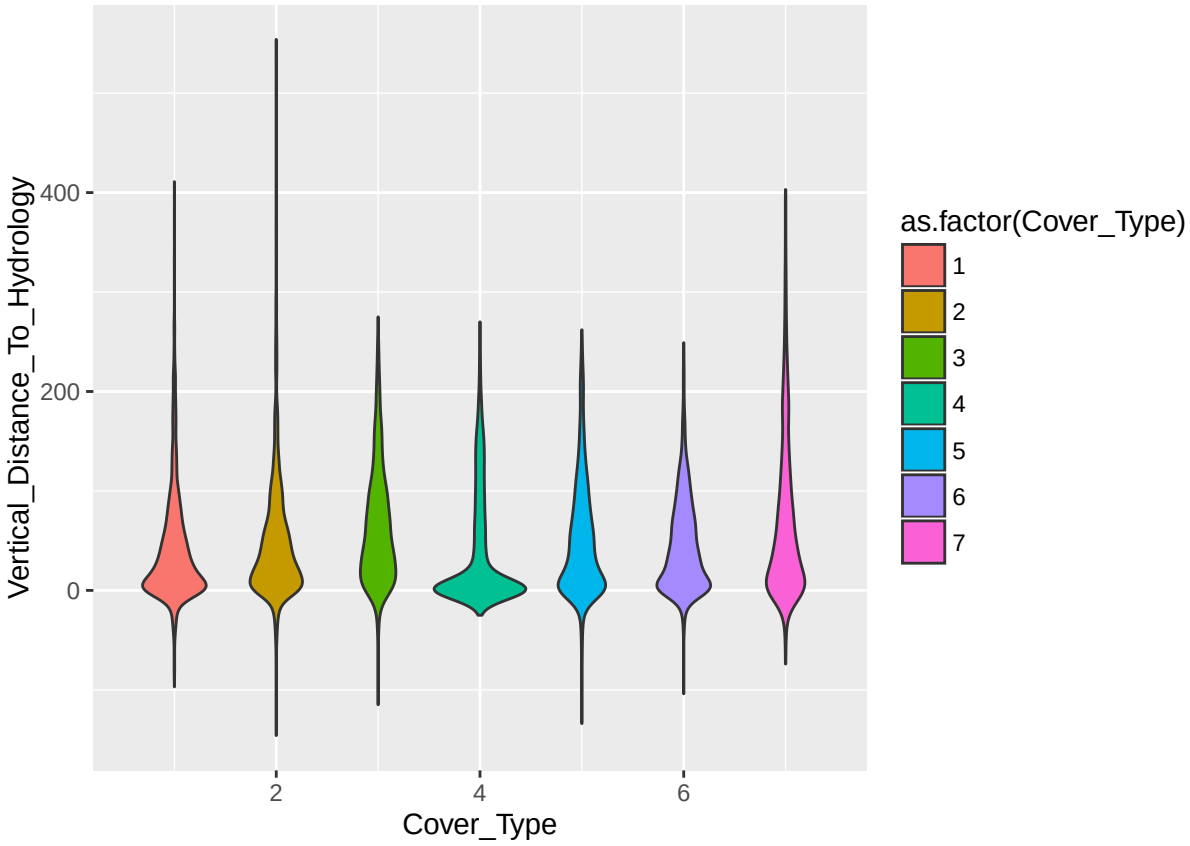
```
for (i in 1:54)
{
print(ggplot(train, aes_string(x=train$Cover_Type, y=xm[i]))+xlab("Cover_Type") + geom_violin(aes(fill=

}
```

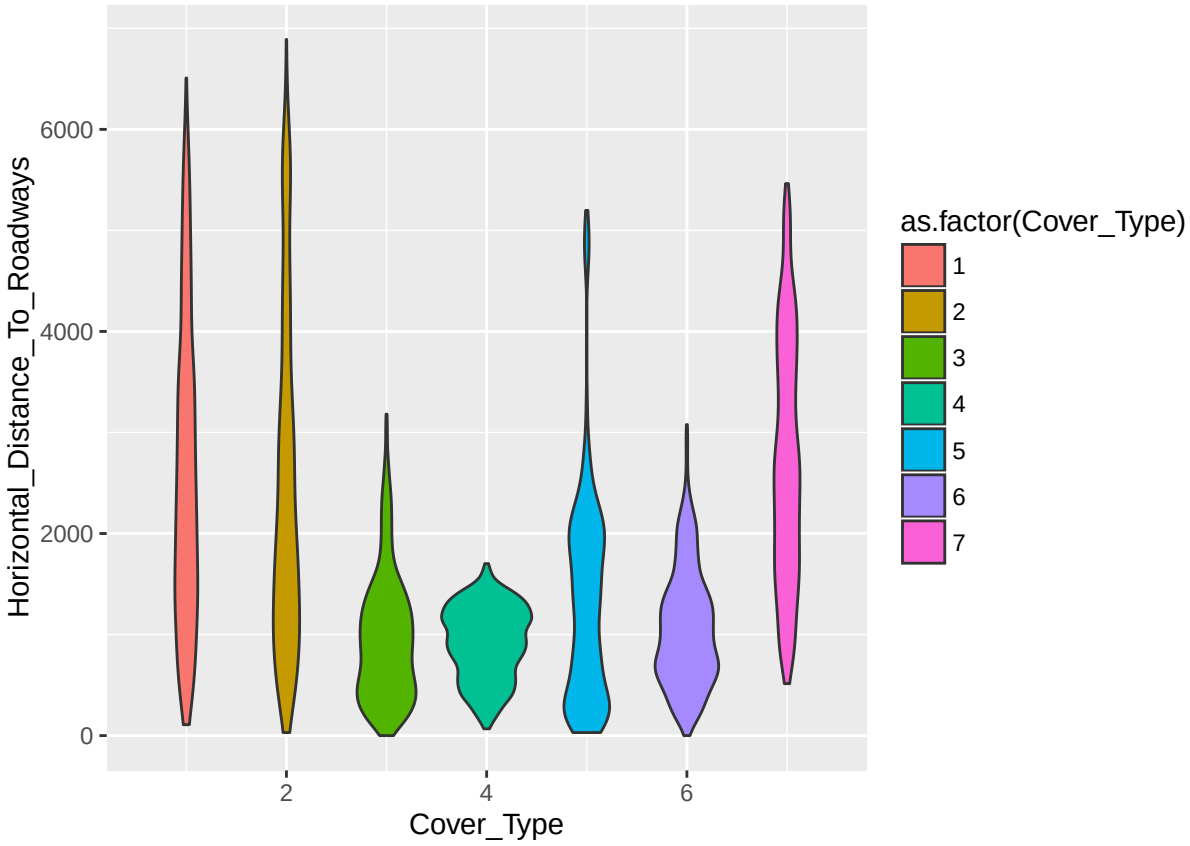

Elevation
3500
3000
2500
2000
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

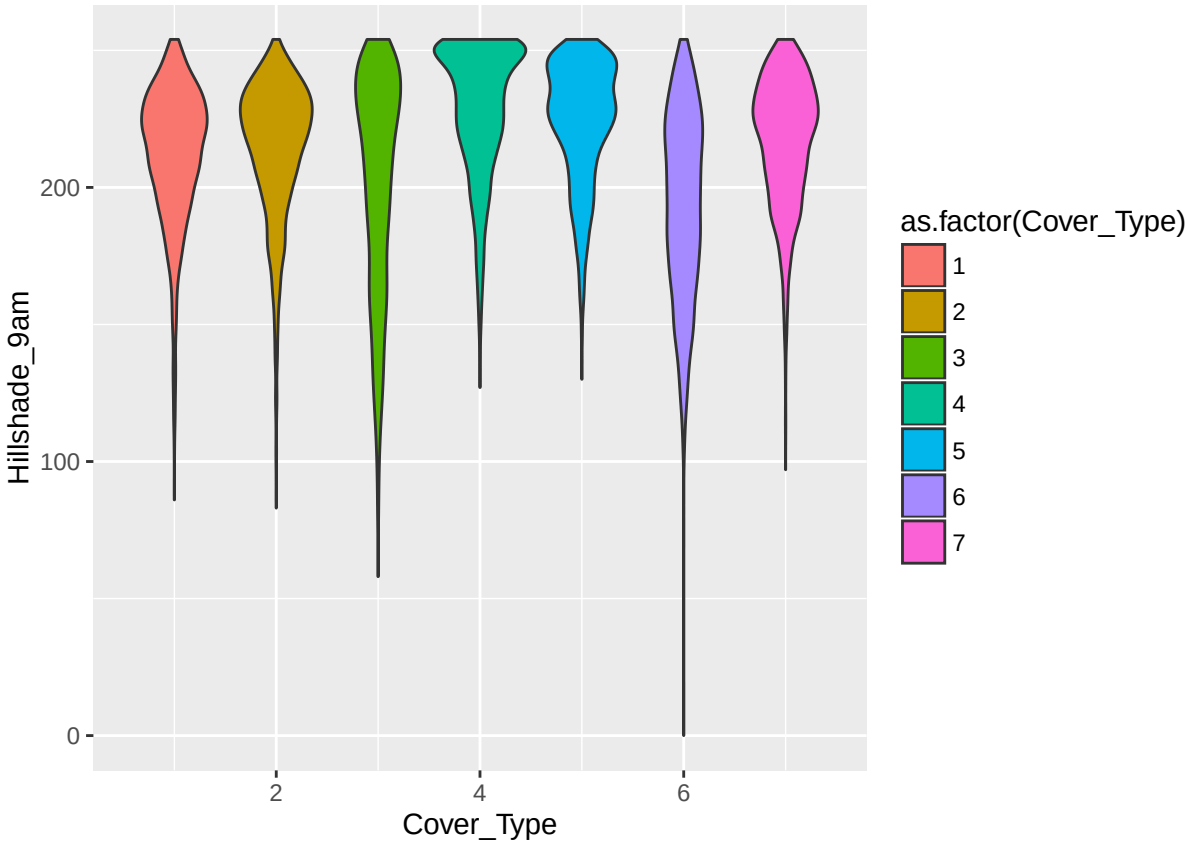

Aspect
300
200
100
0
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

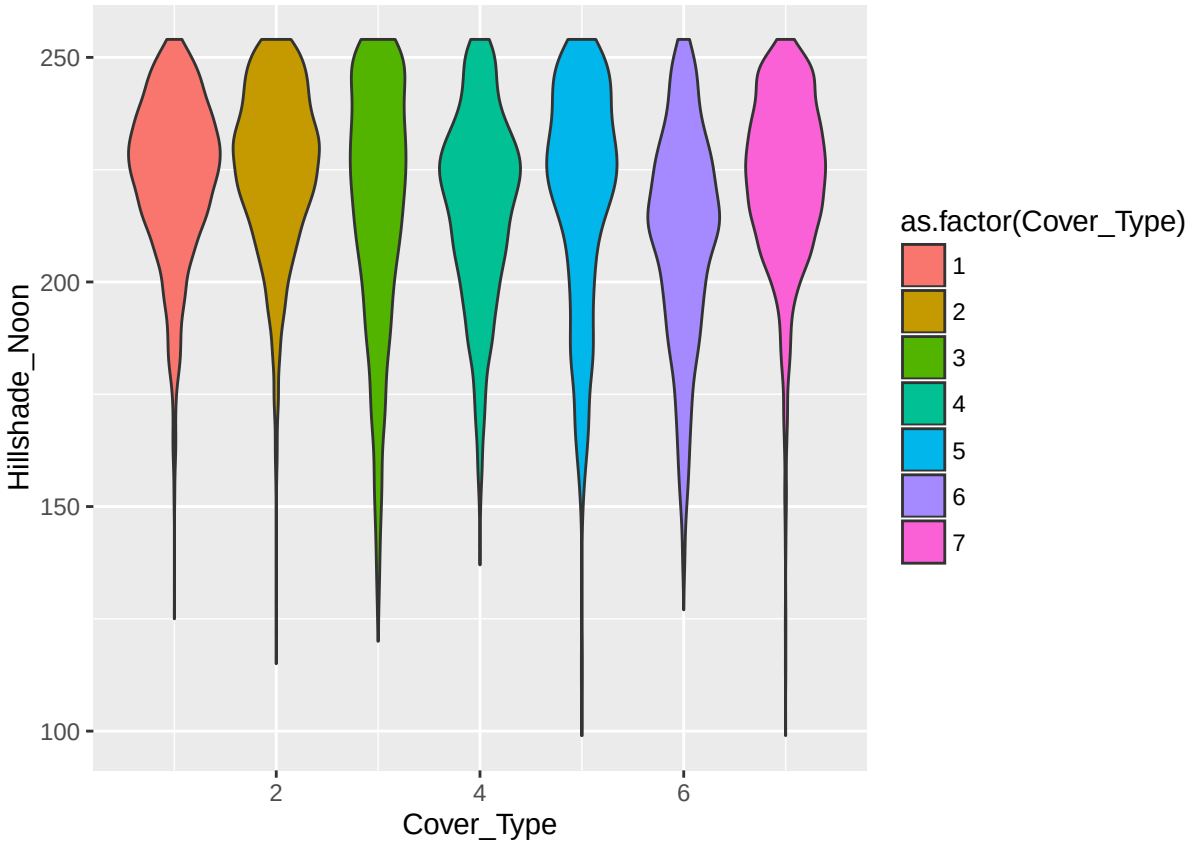

Slope
50
40
30
20
10
0
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

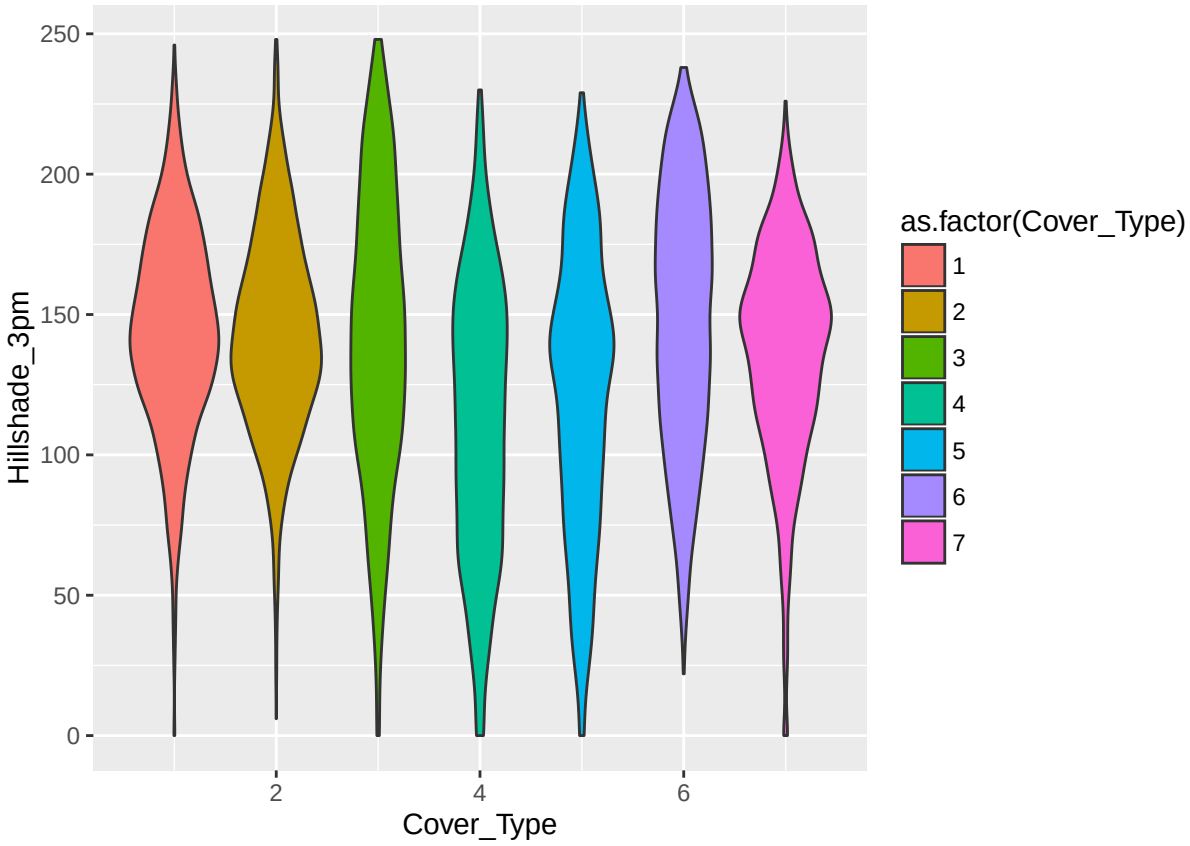

Horizontal_Distance_To_Hydrology
1000
500
0
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

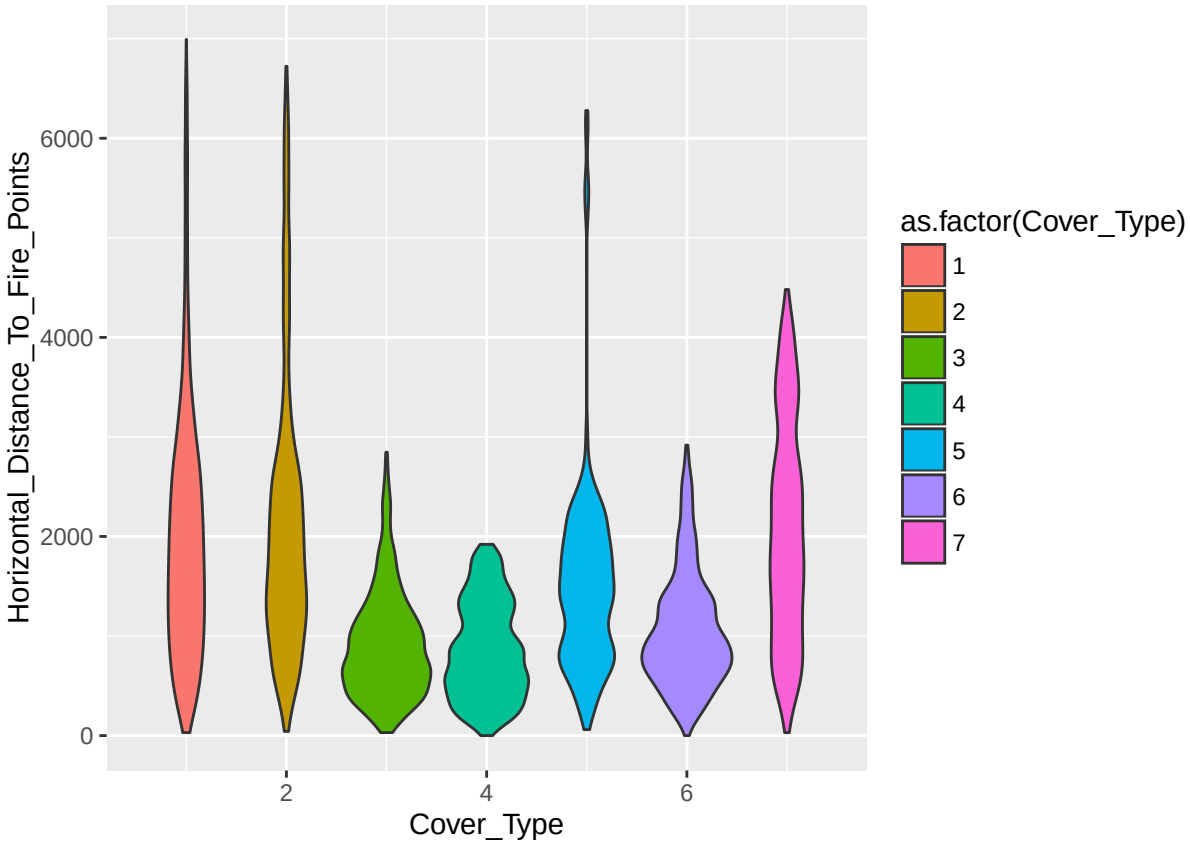

Vertical_Distance_To_Hydrology
400
200
0
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

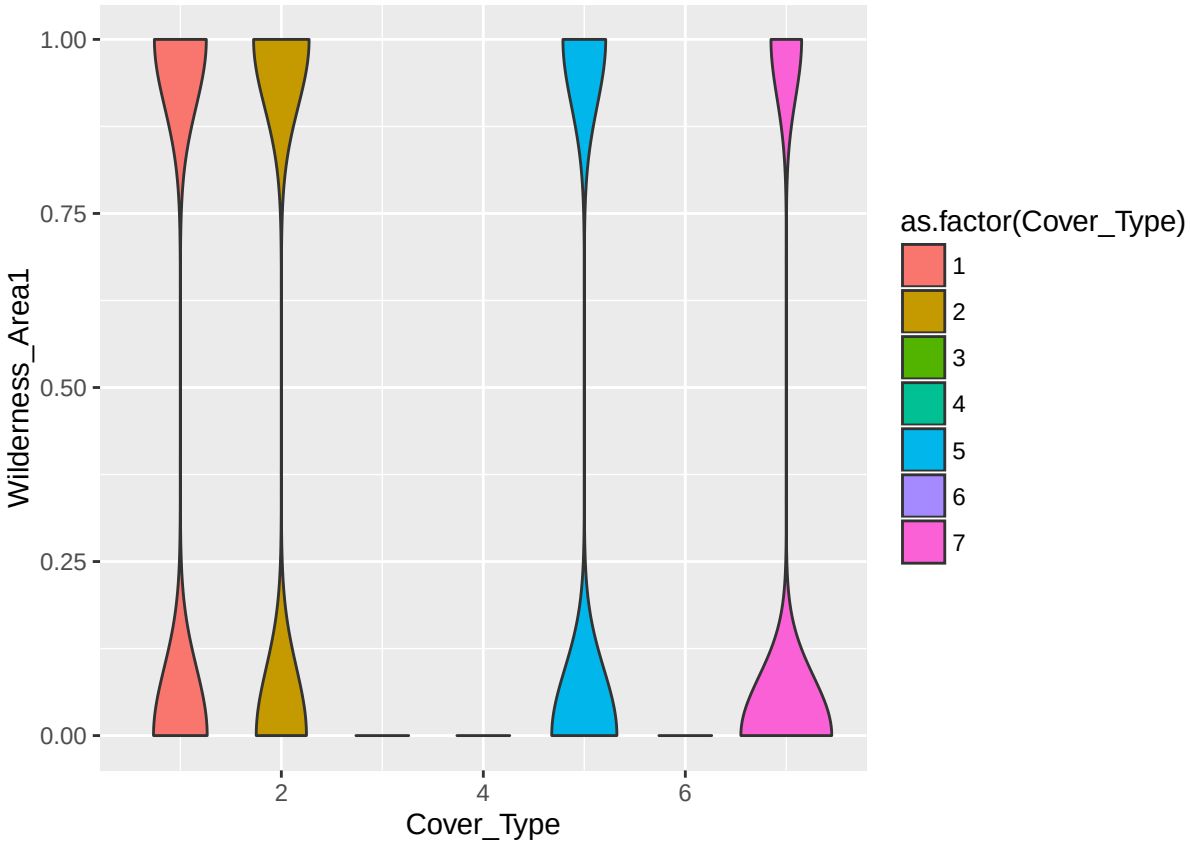

Horizontal_Distance_To_Roadways
6000
4000
2000
0
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

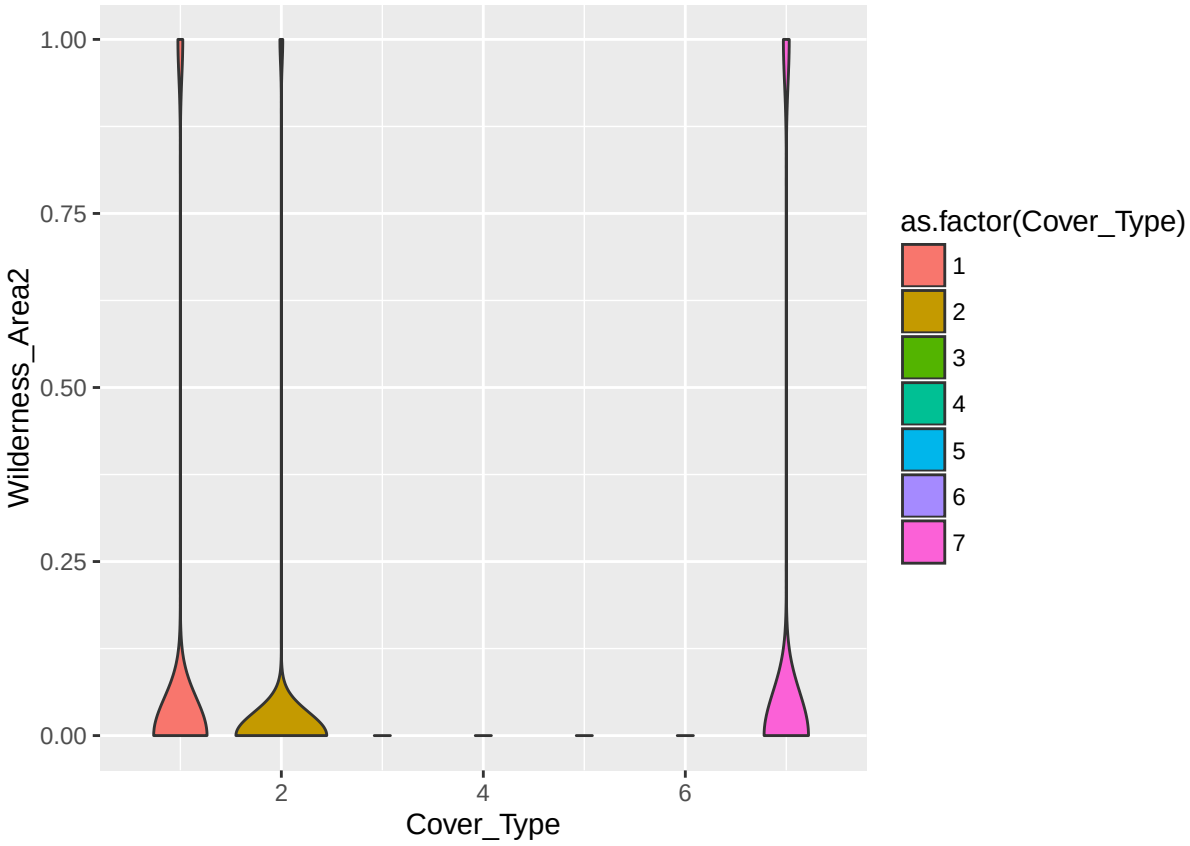

Hillshade_9am
200
100
0
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

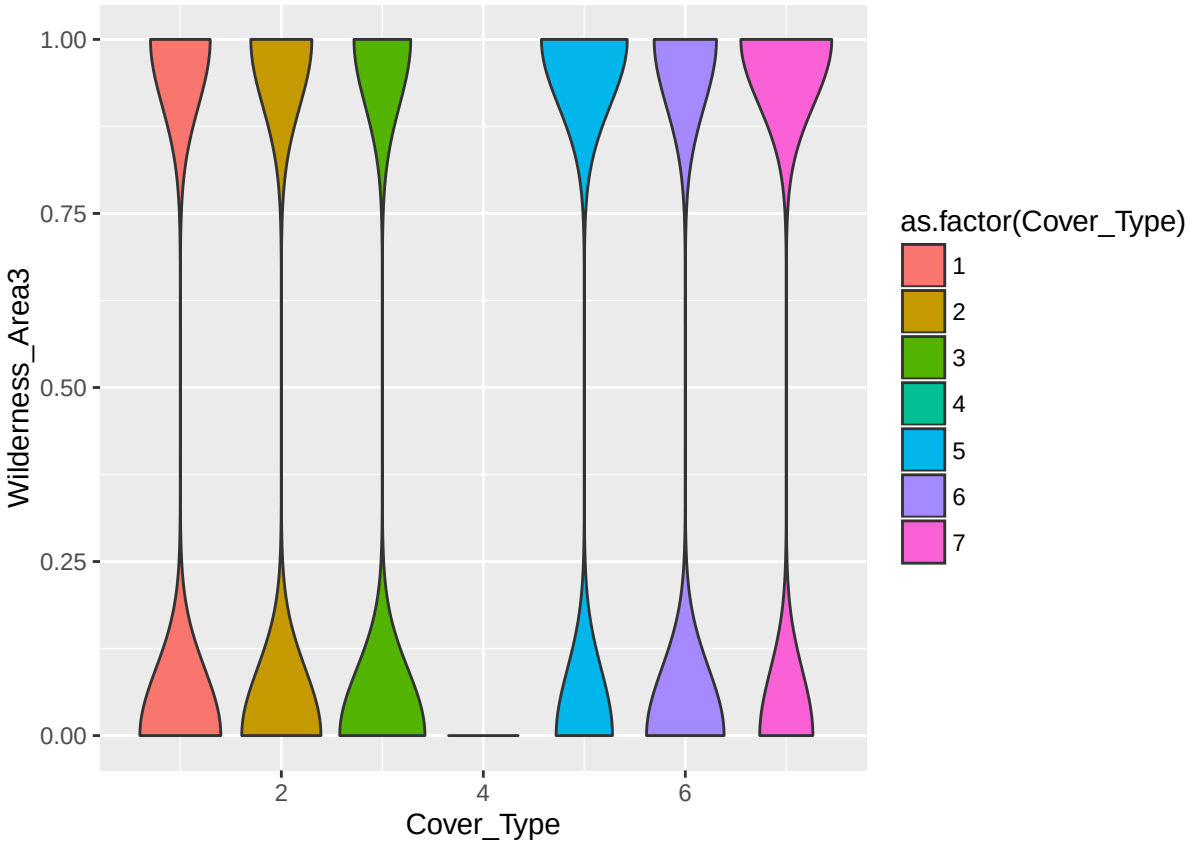

Hillshade_Noon
250
200
150
100
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

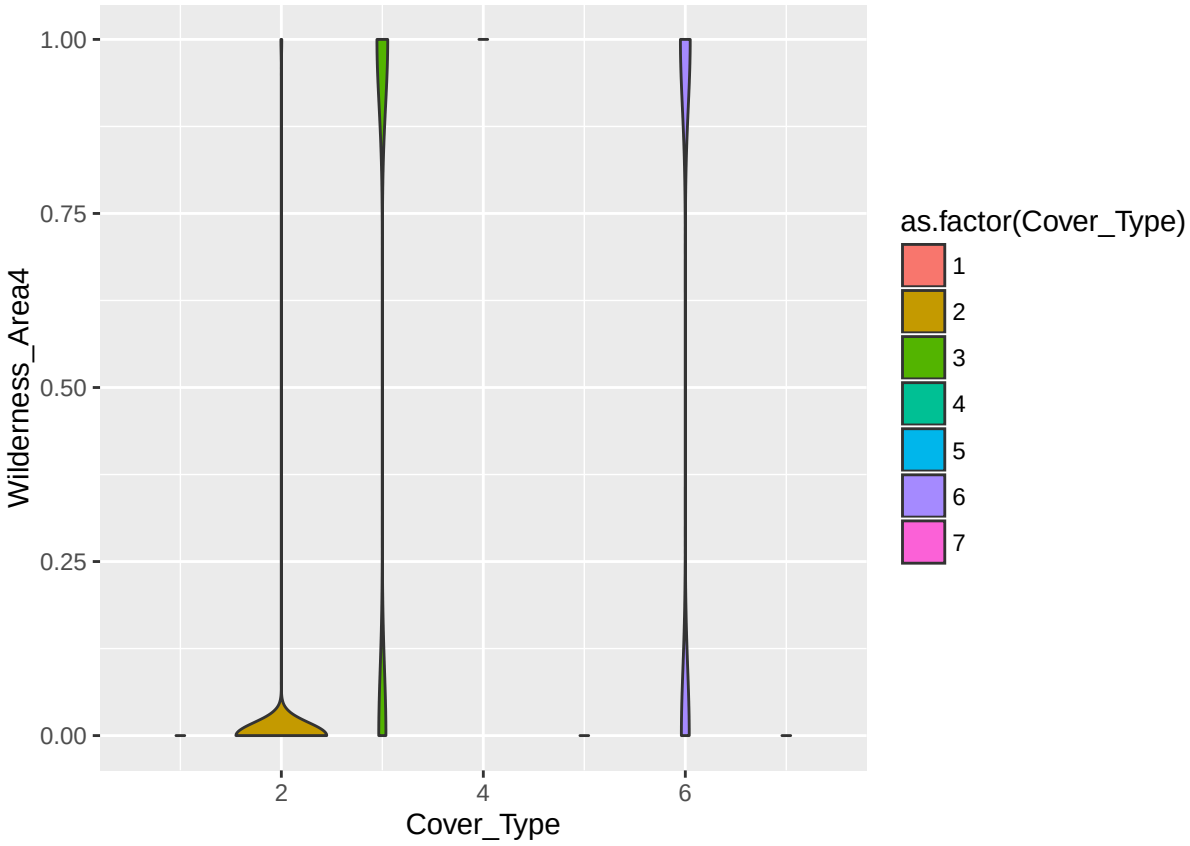

250
200
150
100
50
0
Hillshade_3pm
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

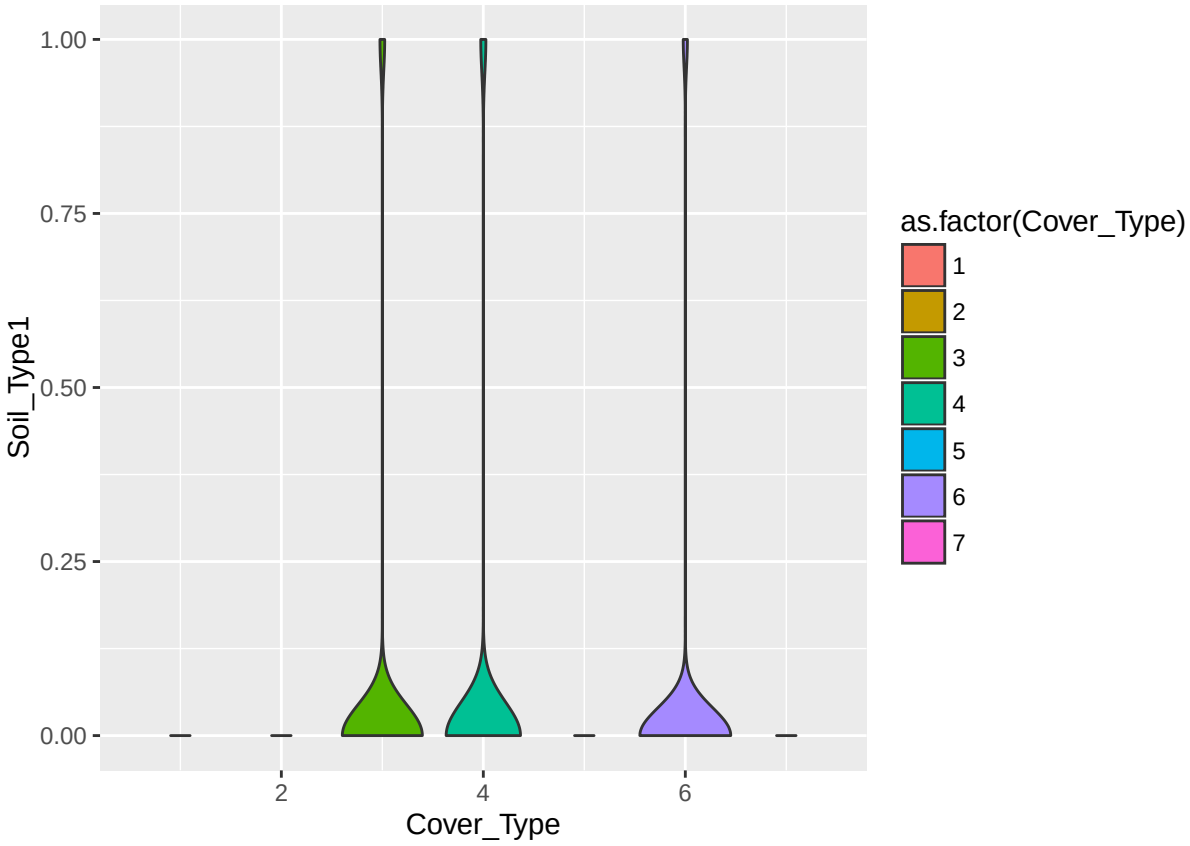

Horizontal_Distance_To_Fire_Points
6000
4000
2000
0
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

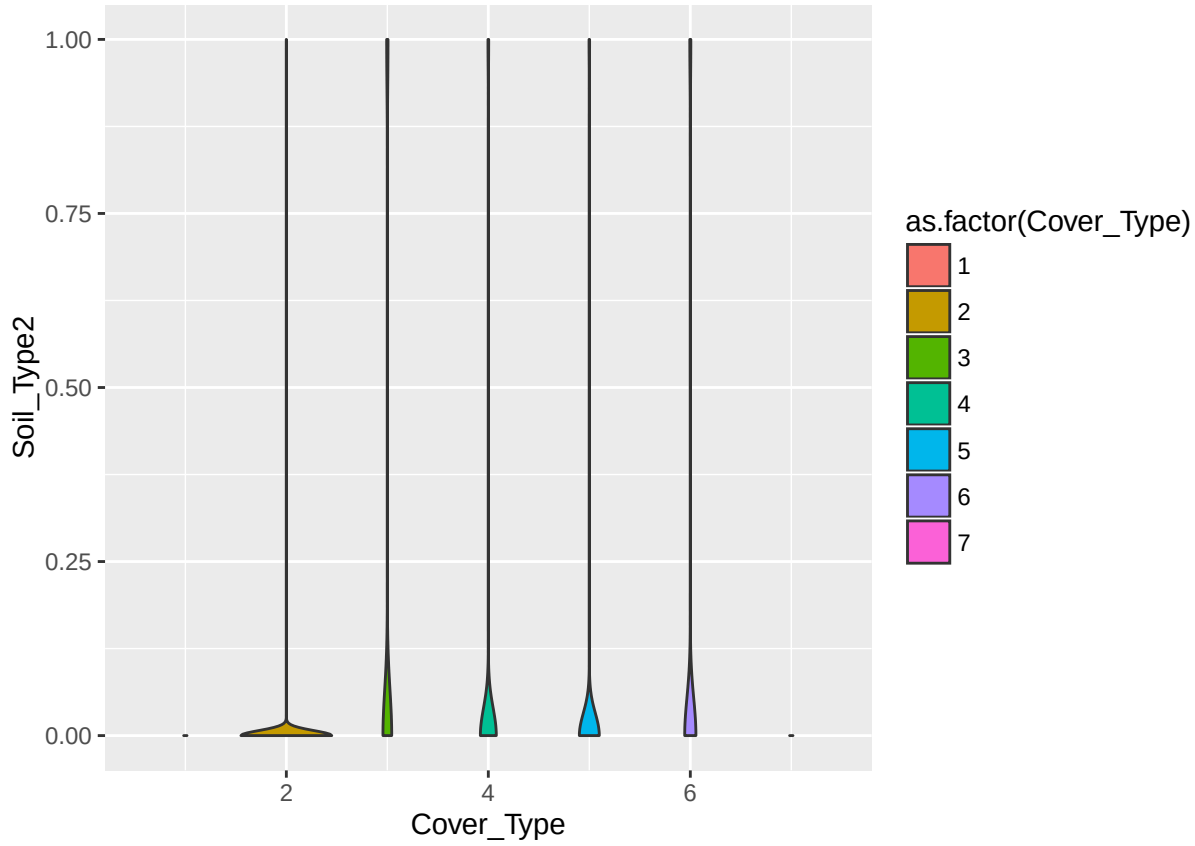

1.00
0.75
0.50
0.25
0.00
Wilderness_Area1
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

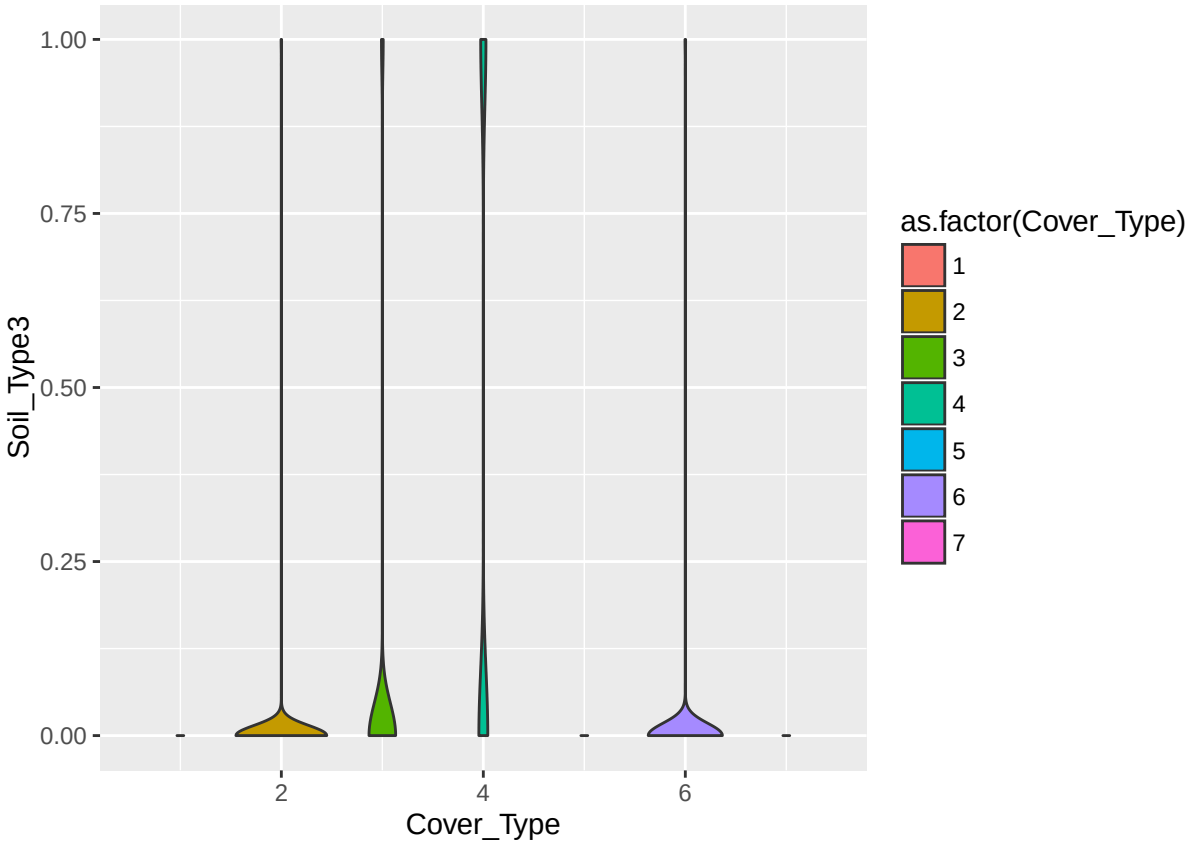

1.00
0.75
0.50
0.25
0.00
Wilderness_Area2
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

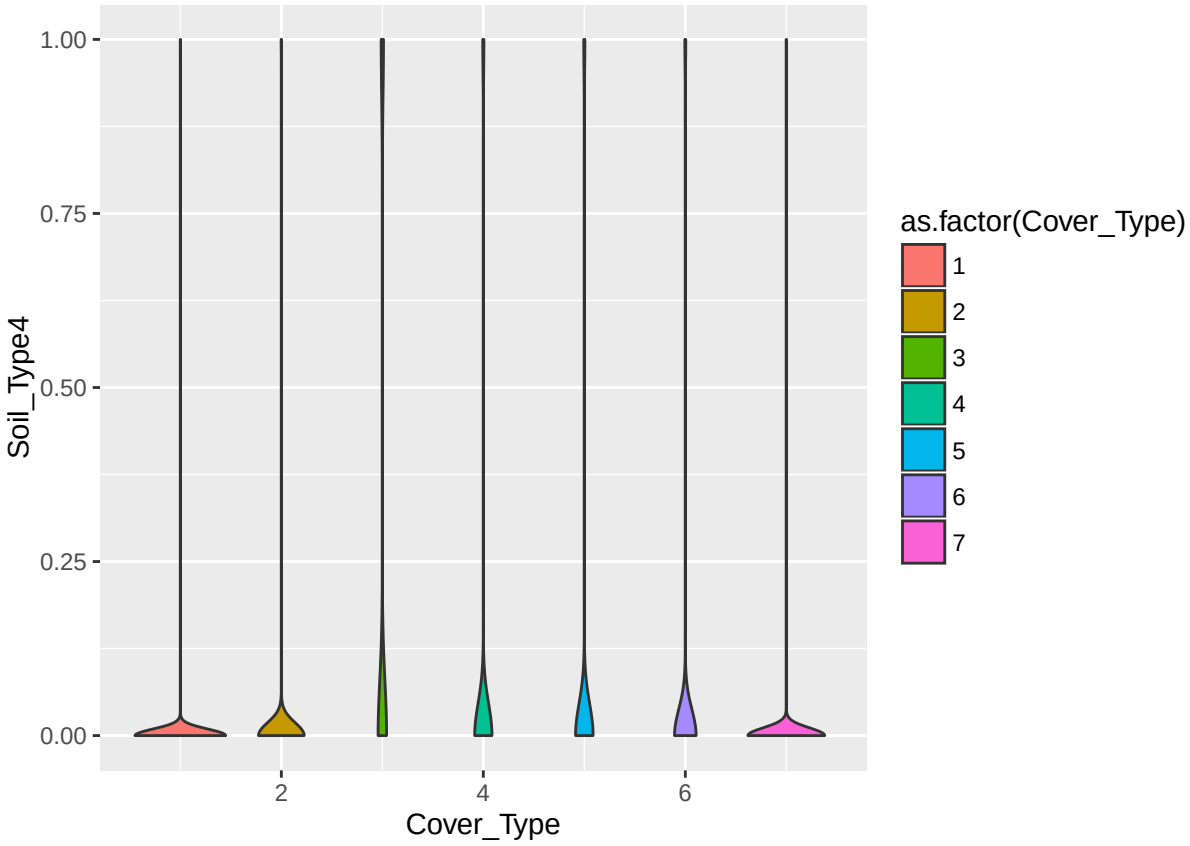

1.00
0.75
0.50
0.25
0.00
Wilderness_Area3
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

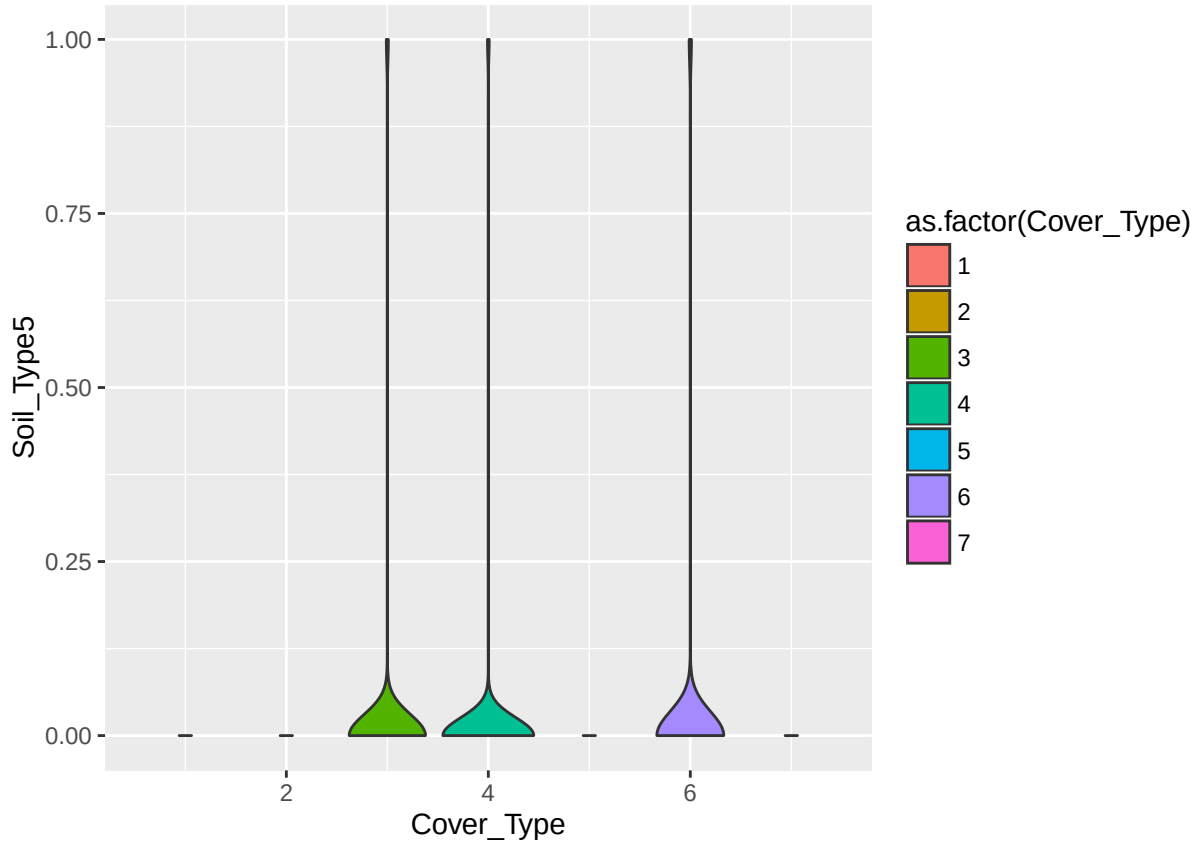

1.00
0.75
0.50
0.25
0.00
Wilderness_Area4
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

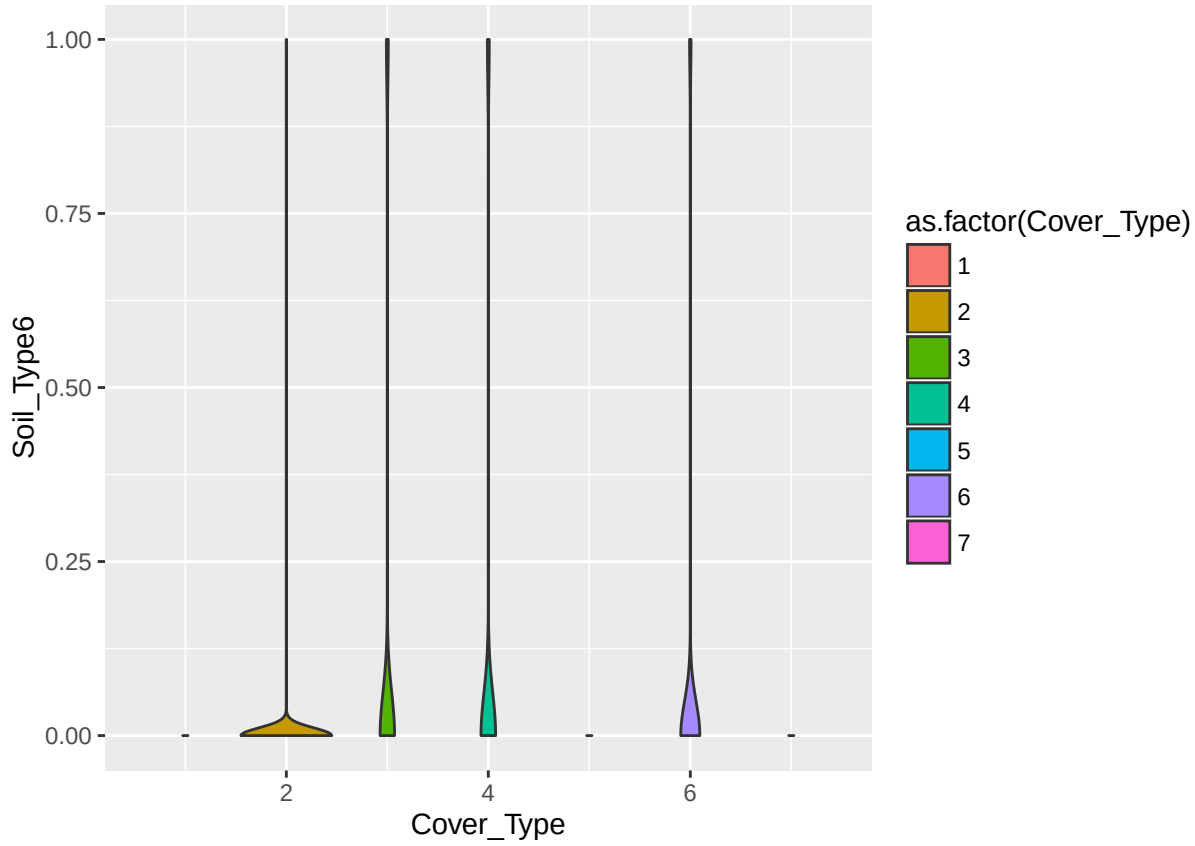

1.00
0.75
0.50
0.25
0.00
Soil_Type1
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

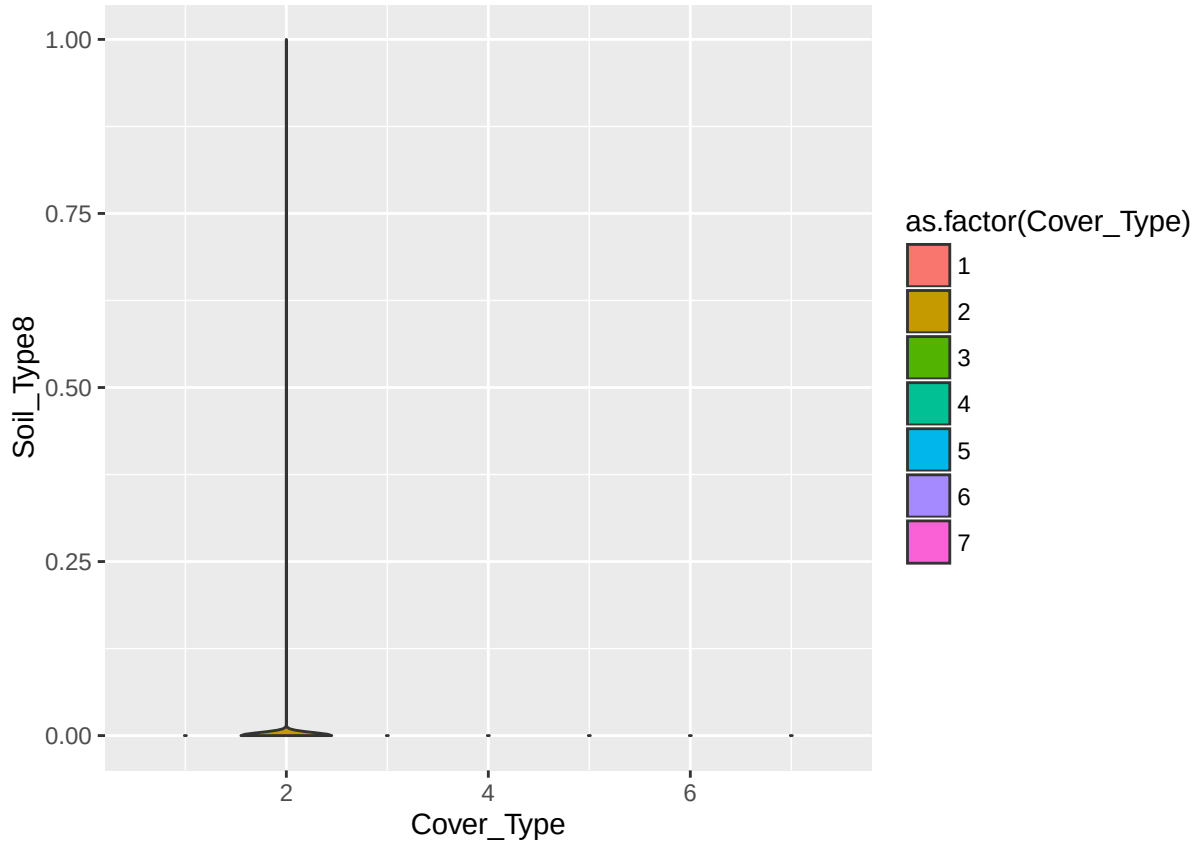

1.00
0.75
0.50
0.25
0.00
Soil_Type2
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

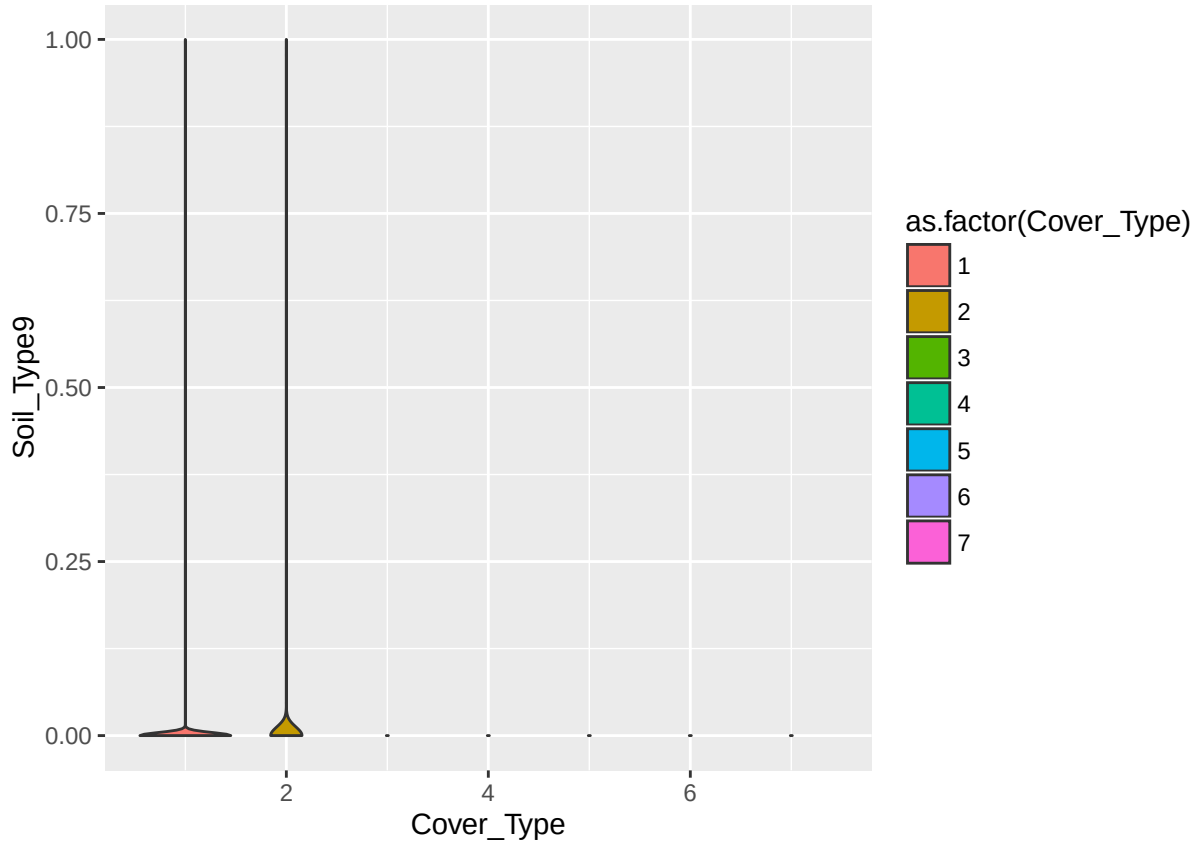

1.00
0.75
0.50
0.25
0.00
Soil_Type3
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

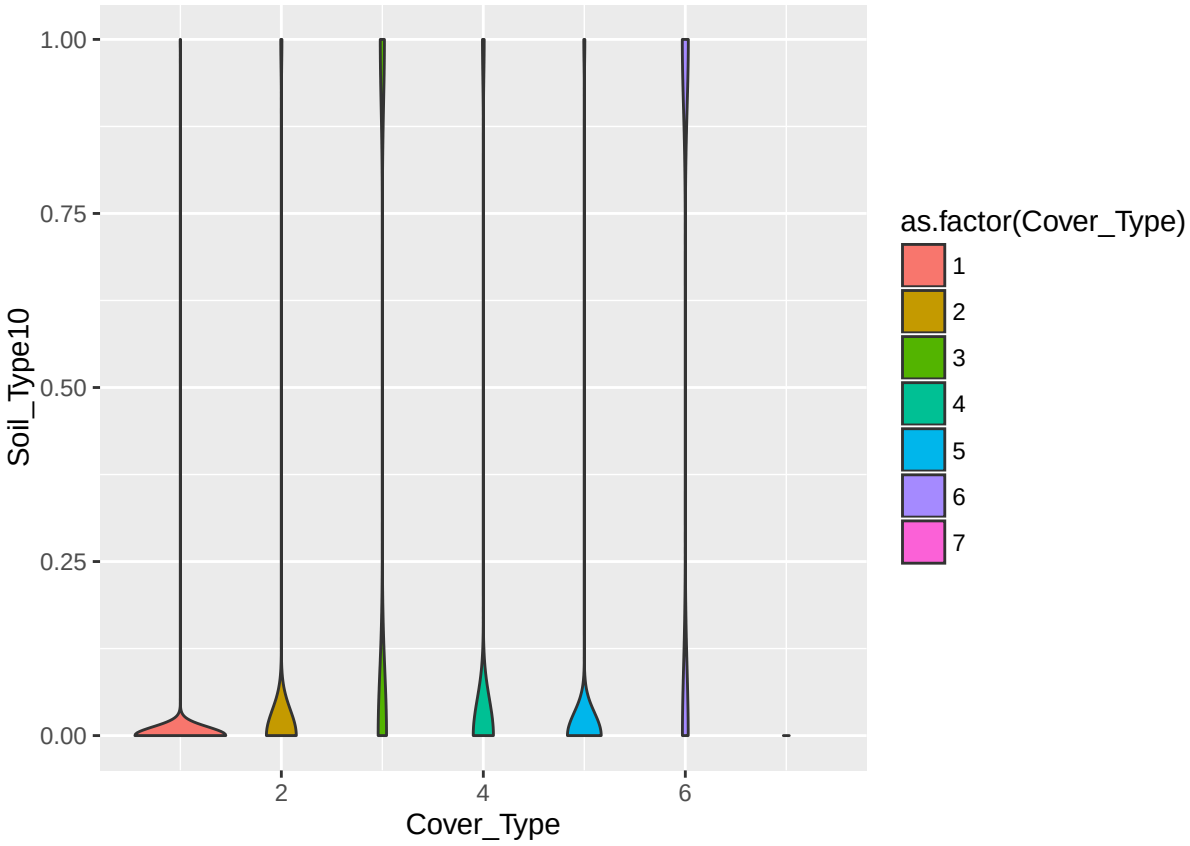

1.00
0.75
0.50
0.25
0.00
Soil_Type4
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

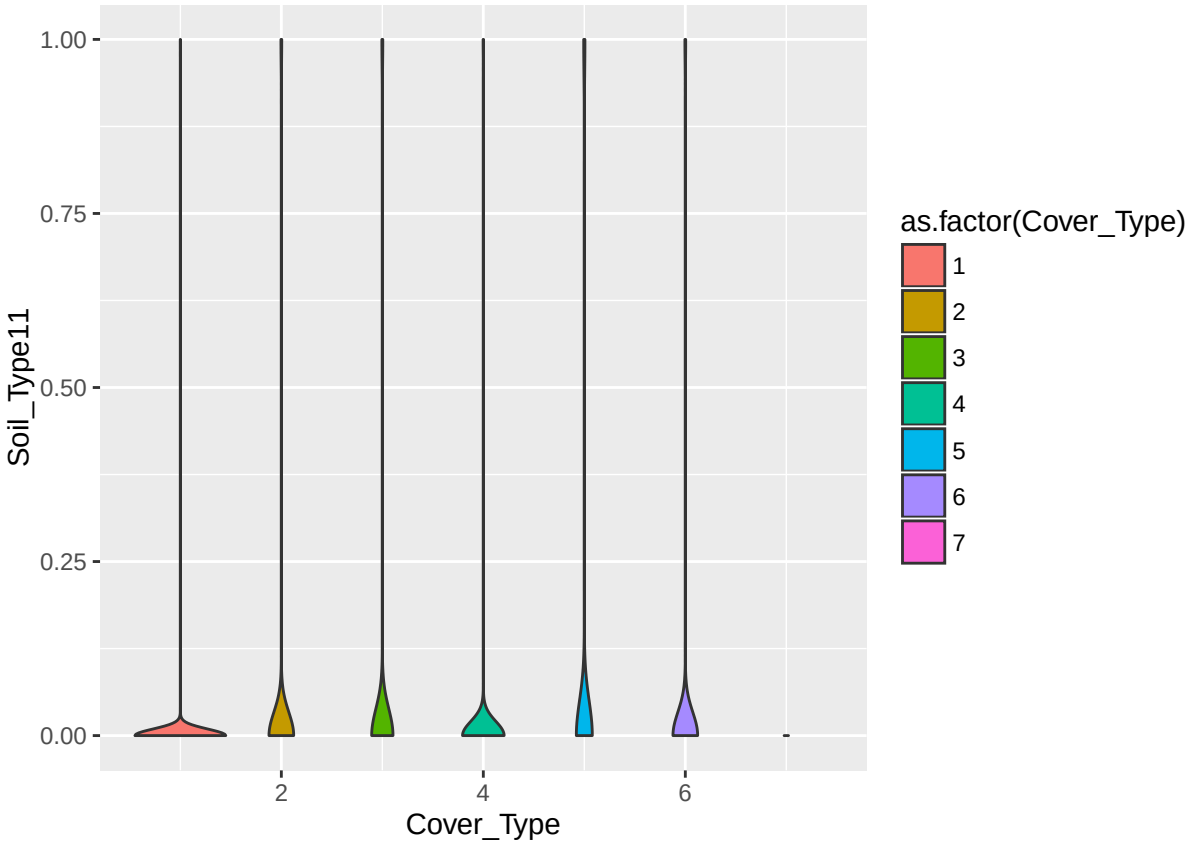

1.00
0.75
0.50
0.25
0.00
Soil_Type5
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

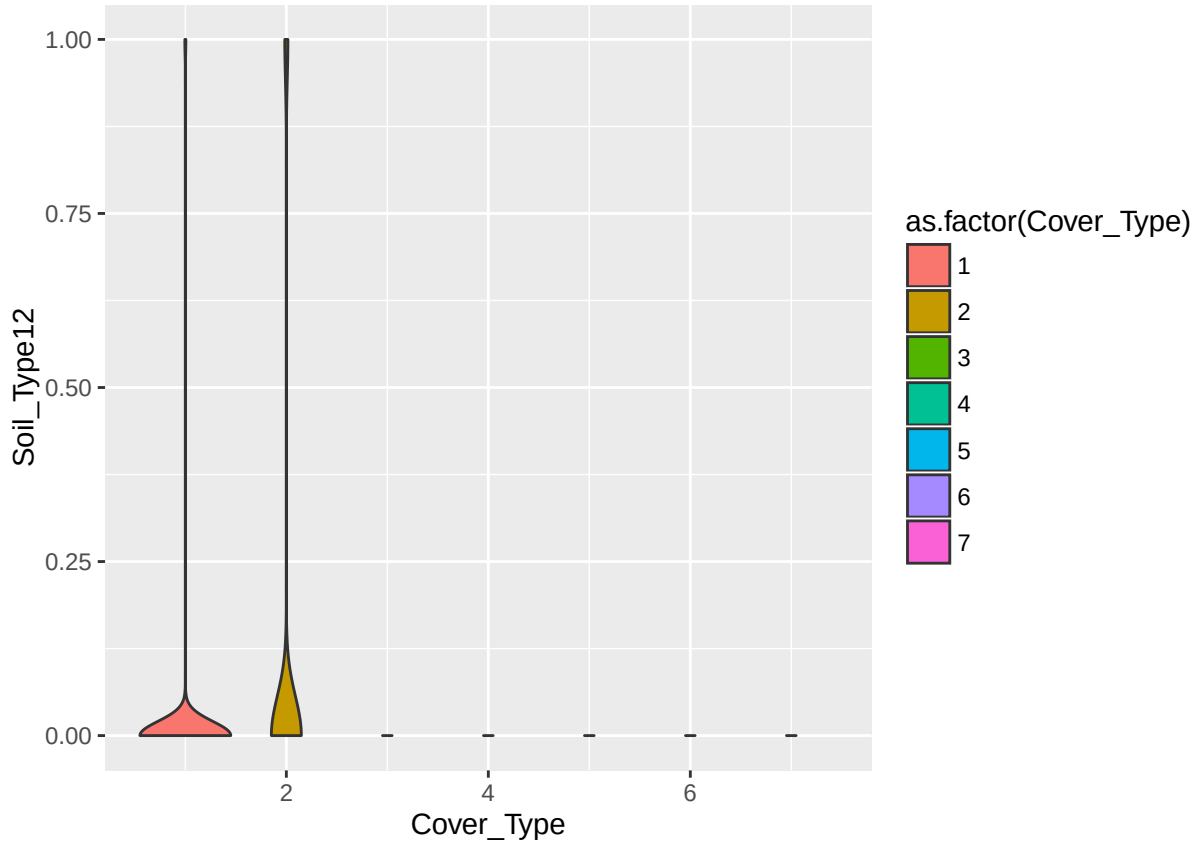

1.00
0.75
0.50
0.25
0.00
Soil_Type6
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

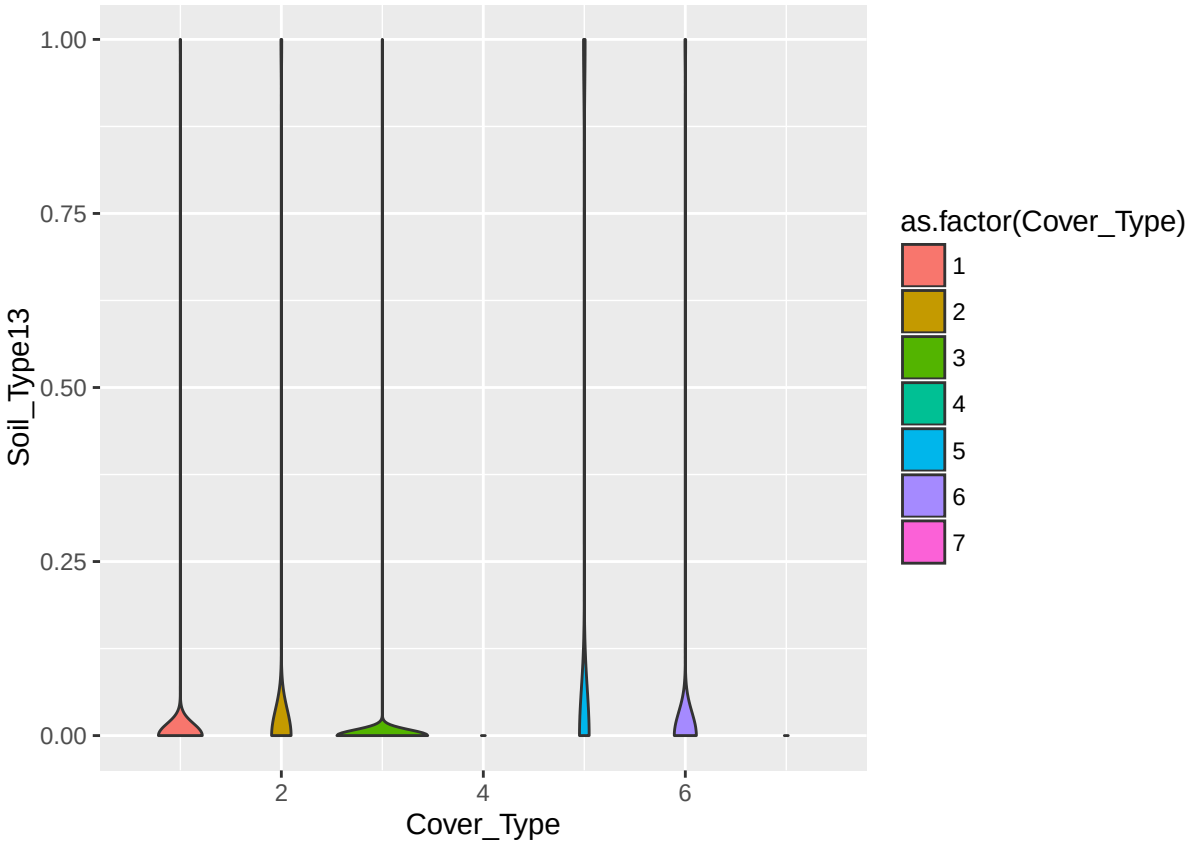

1.00
0.75
0.50
0.25
0.00
Soil_Type8
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

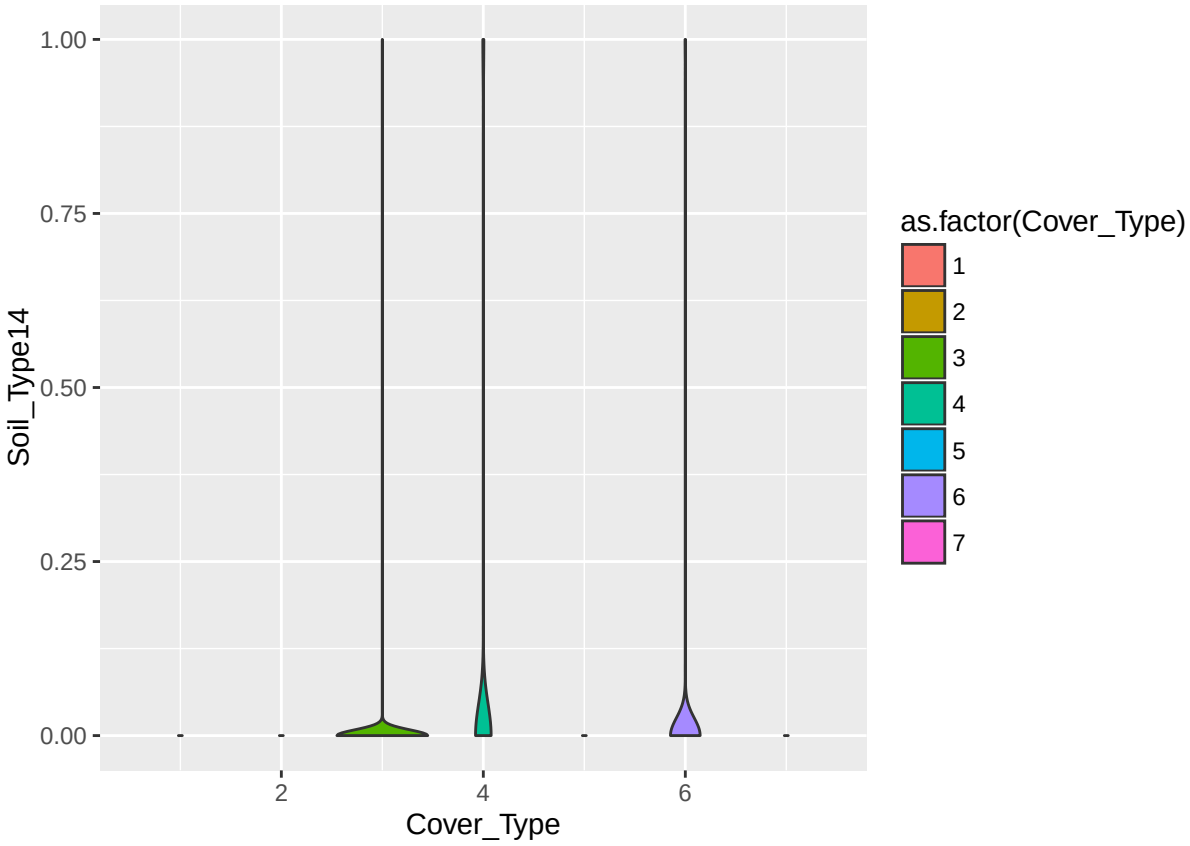

1.00
0.75
0.50
0.25
0.00
Soil_Type9
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

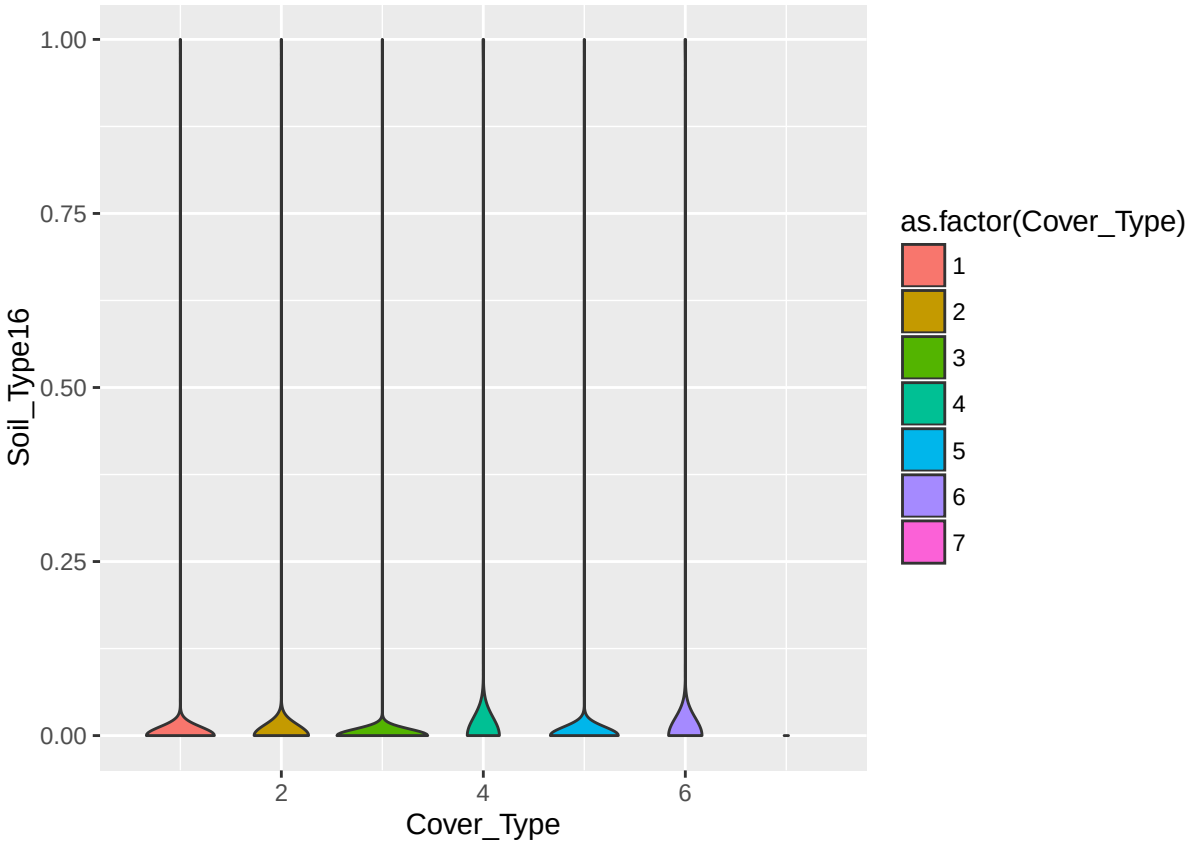

1.00
0.75
0.50
0.25
0.00
Soil_Type10
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

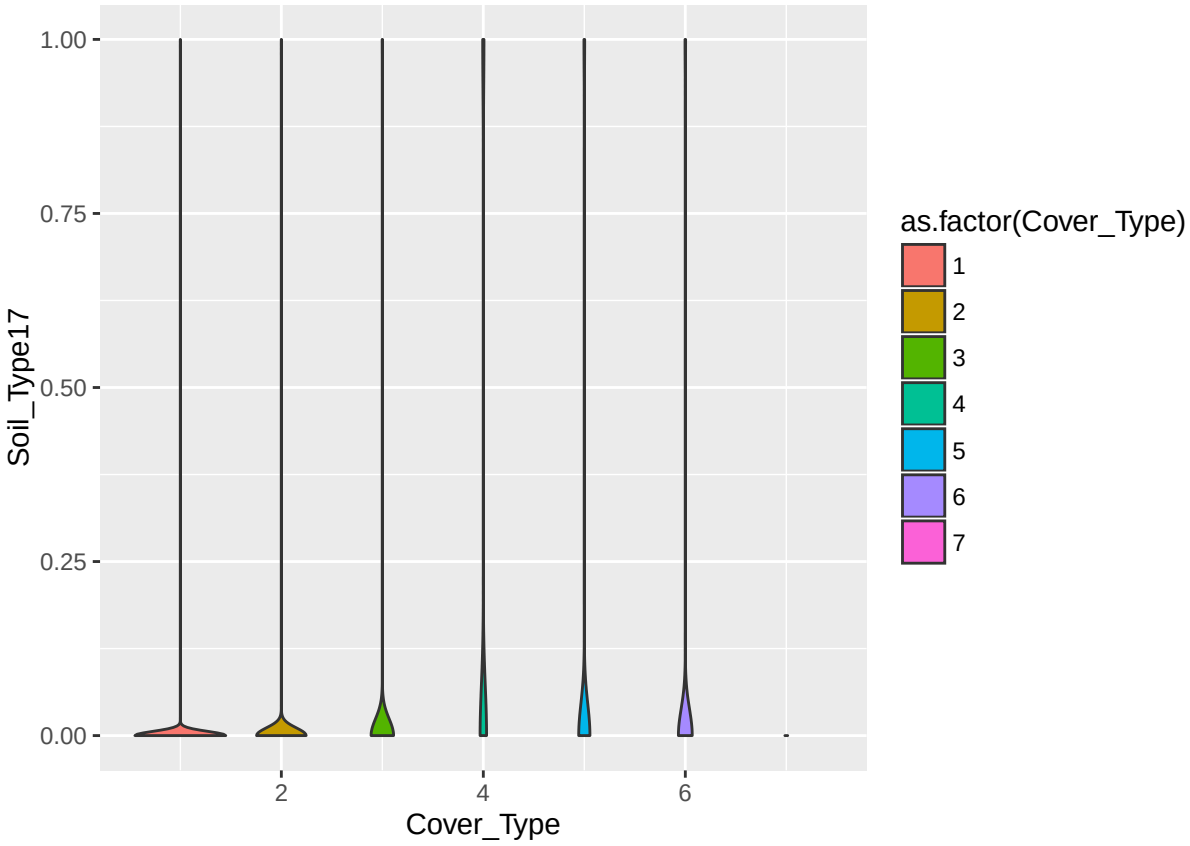

1.00
0.75
0.50
0.25
0.00
Soil_Type11
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

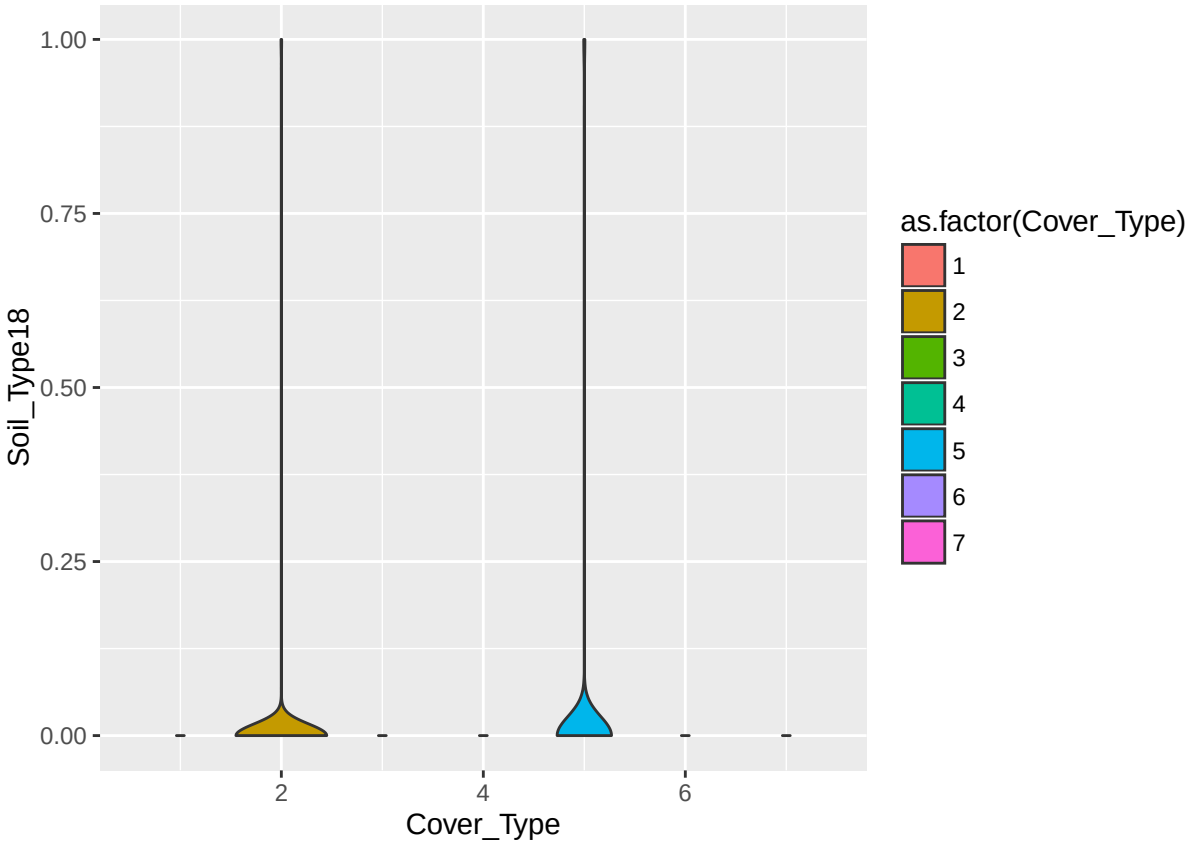

1.00
0.75
0.50
0.25
0.00
Soil_Type12
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

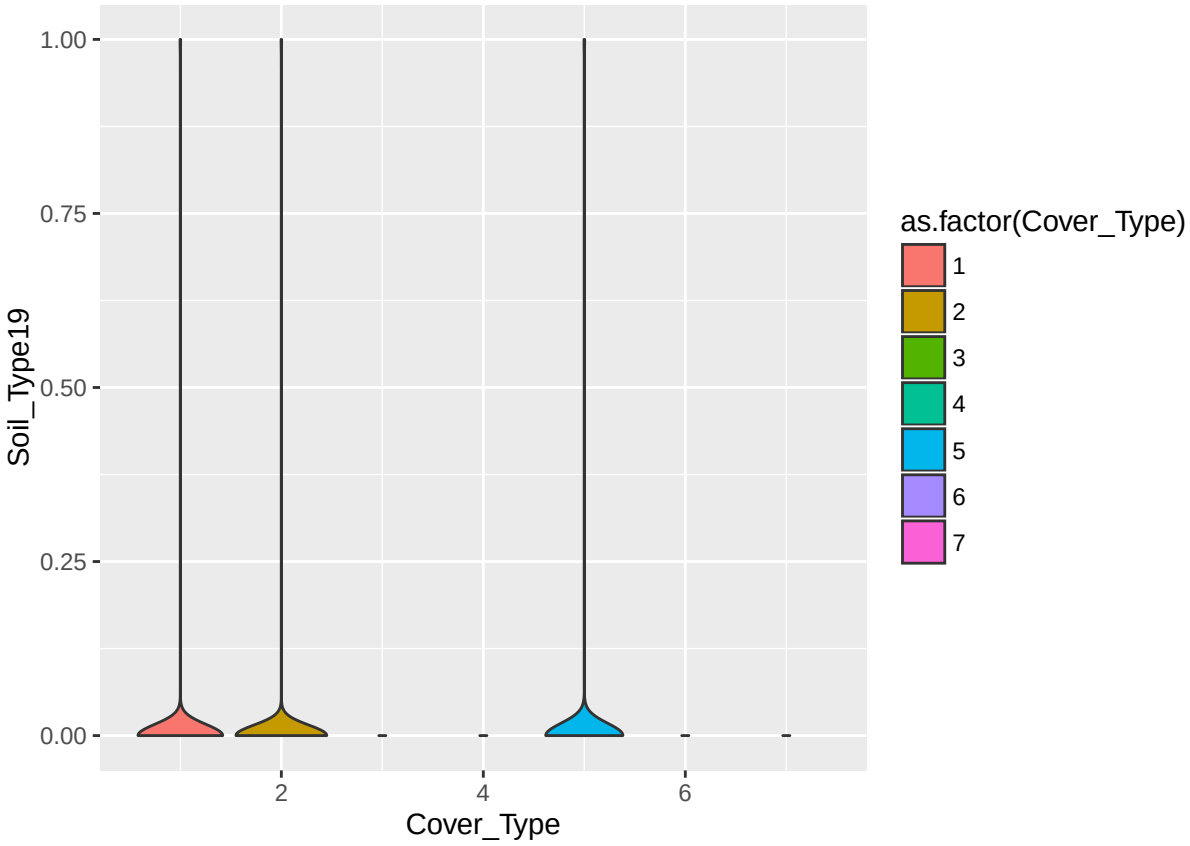

1.00
0.75
0.50
0.25
0.00
Soil_Type13
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

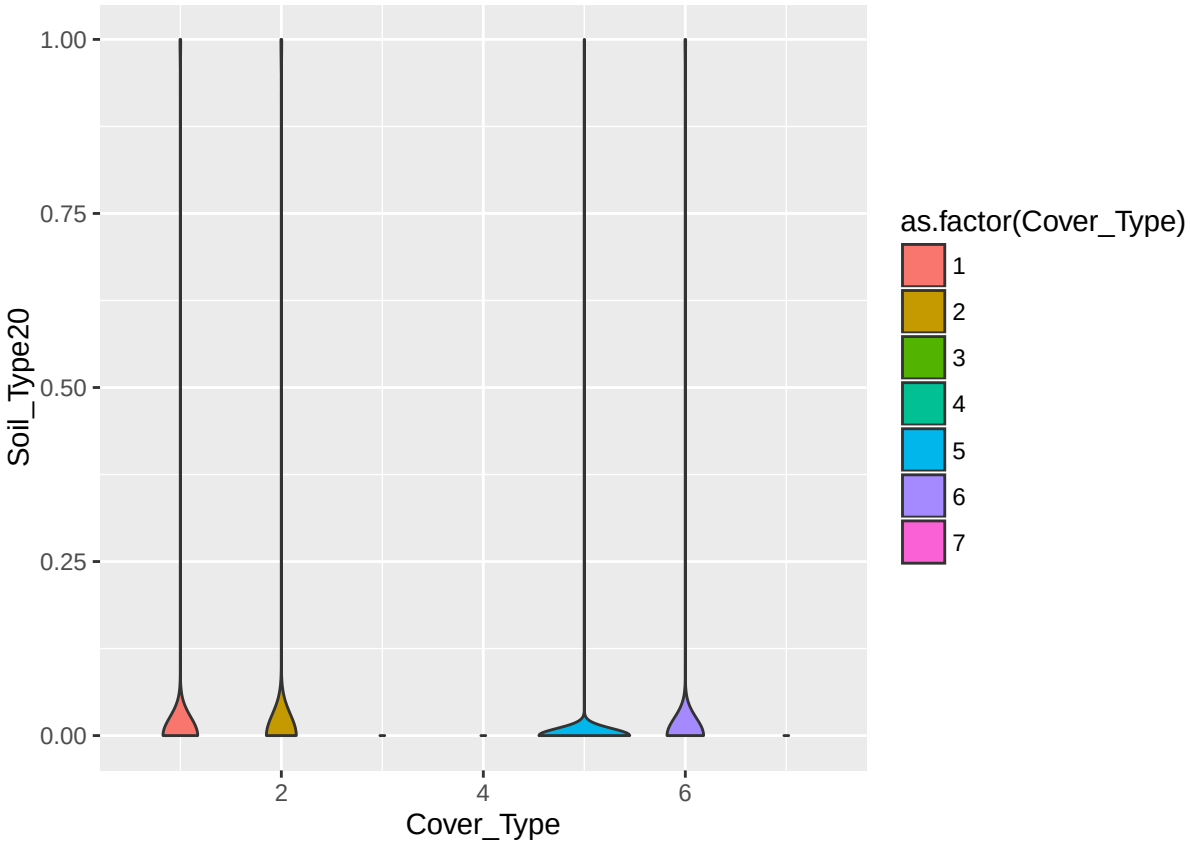

1.00
0.75
0.50
0.25
0.00
Soil_Type14
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

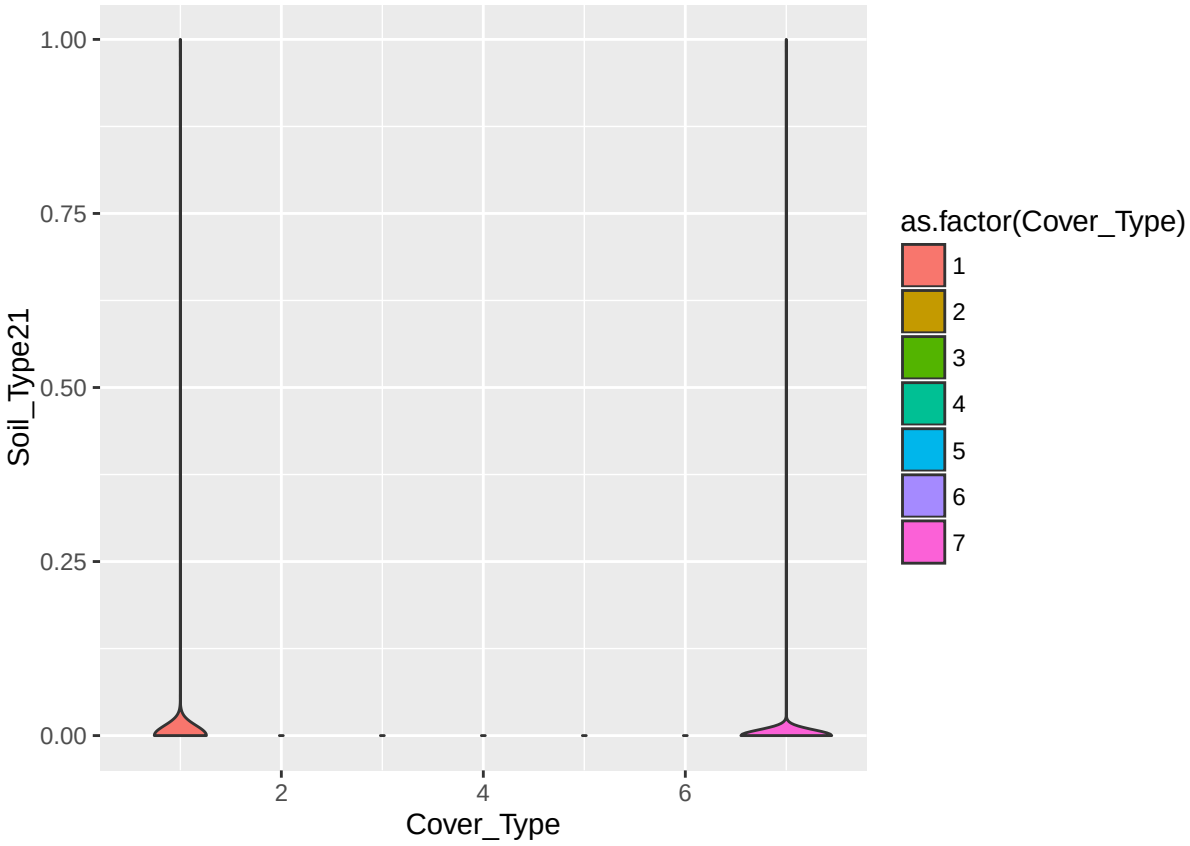

1.00
0.75
0.50
0.25
0.00
Soil_Type16
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

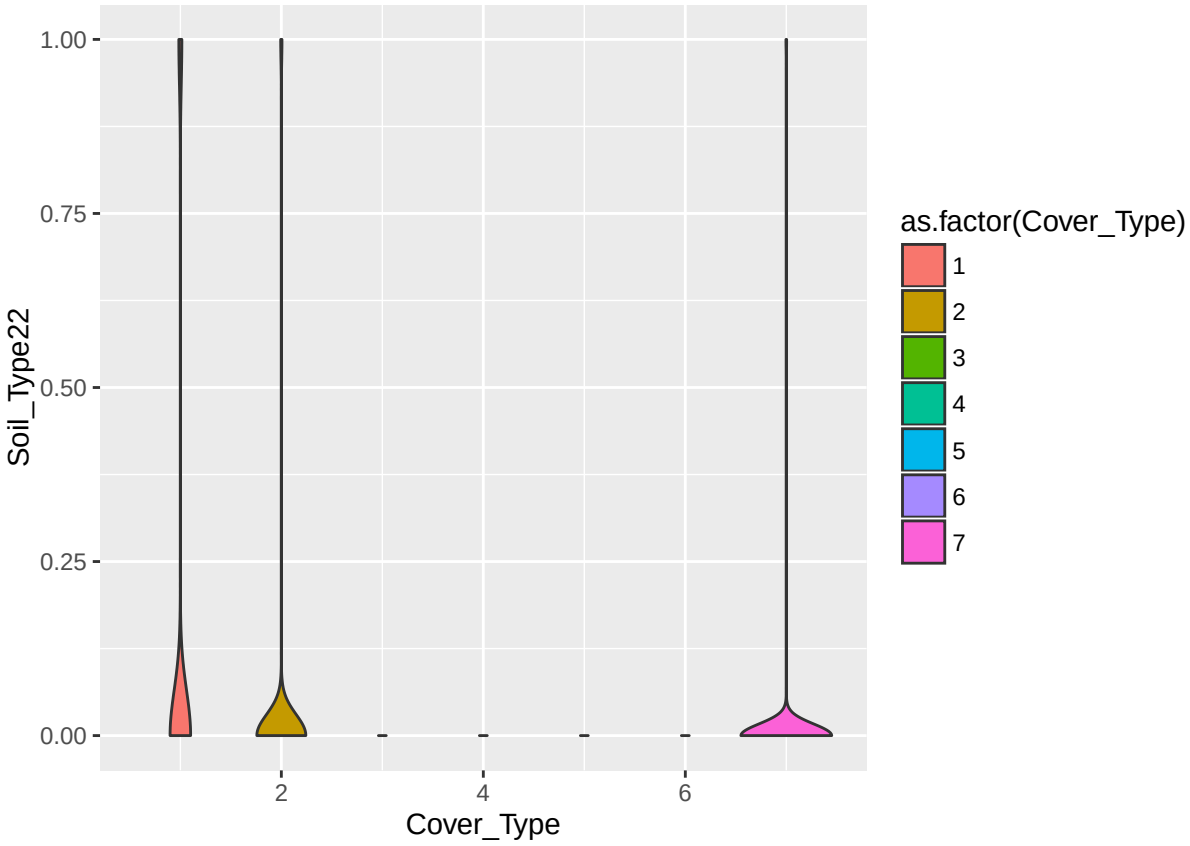

1.00
0.75
0.50
0.25
0.00
Soil_Type17
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

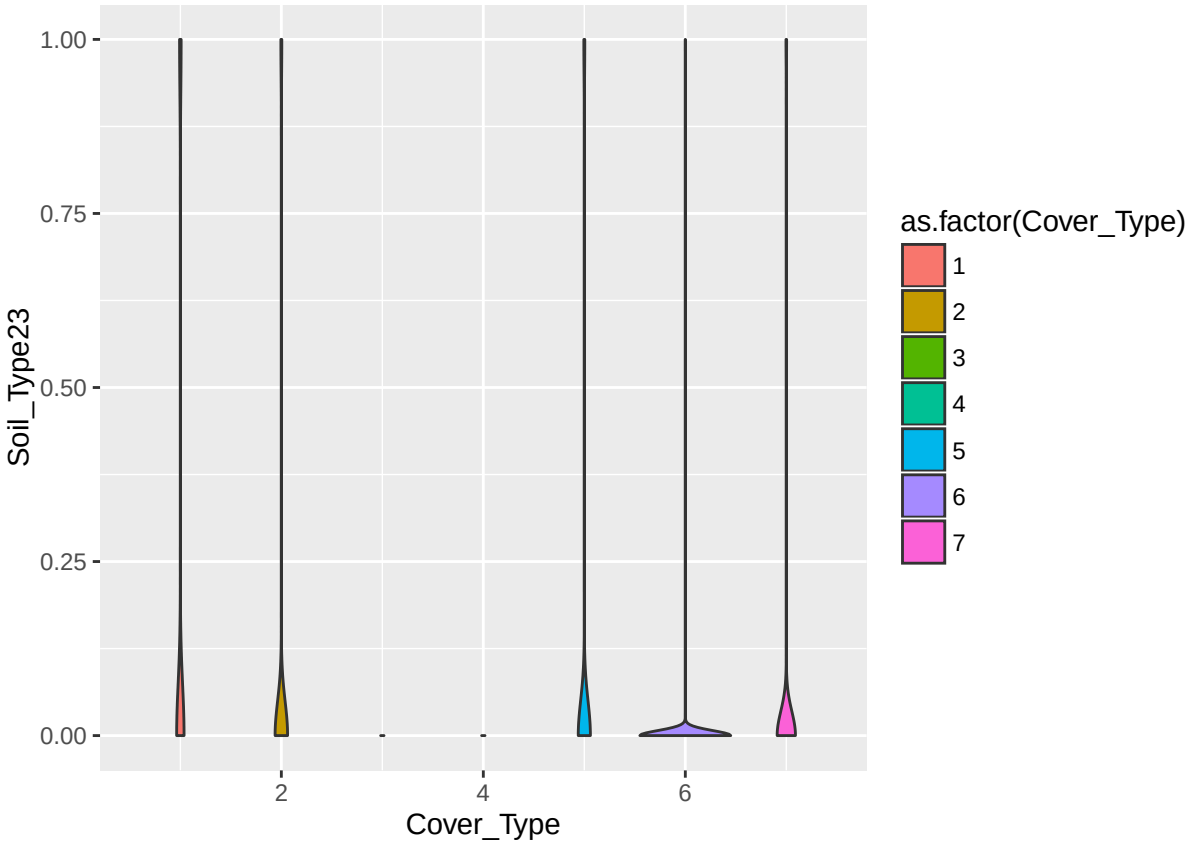

1.00
0.75
0.50
0.25
0.00
Soil_Type18
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

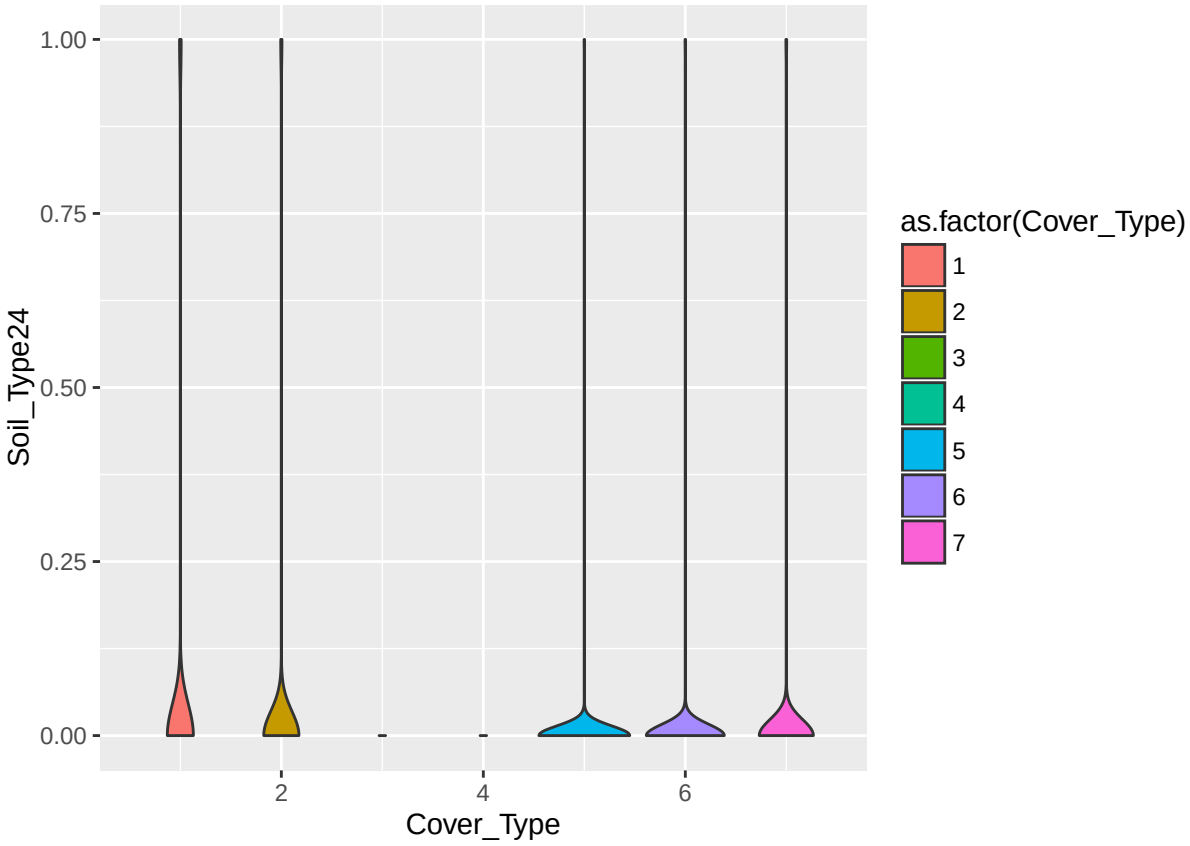

1.00
0.75
0.50
0.25
0.00
Soil_Type19
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

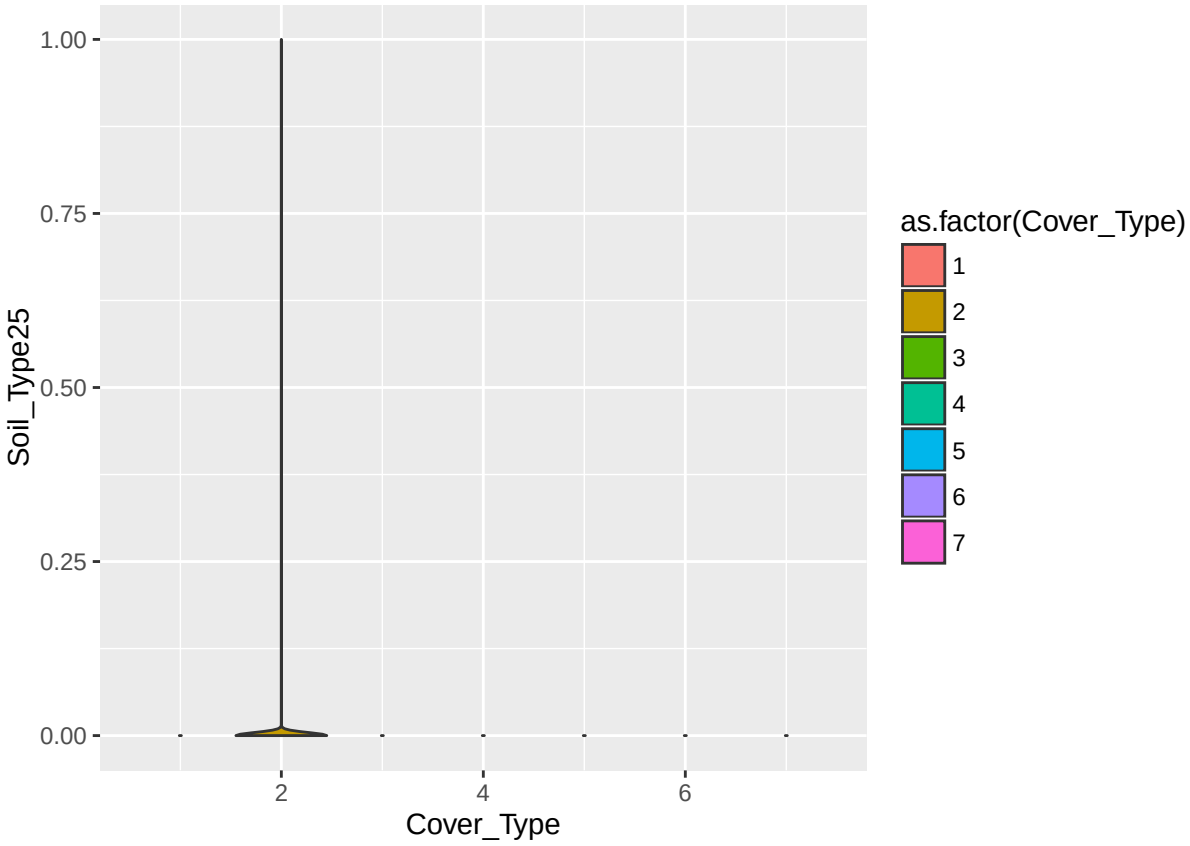

1.00
0.75
0.50
0.25
0.00
Soil_Type20
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

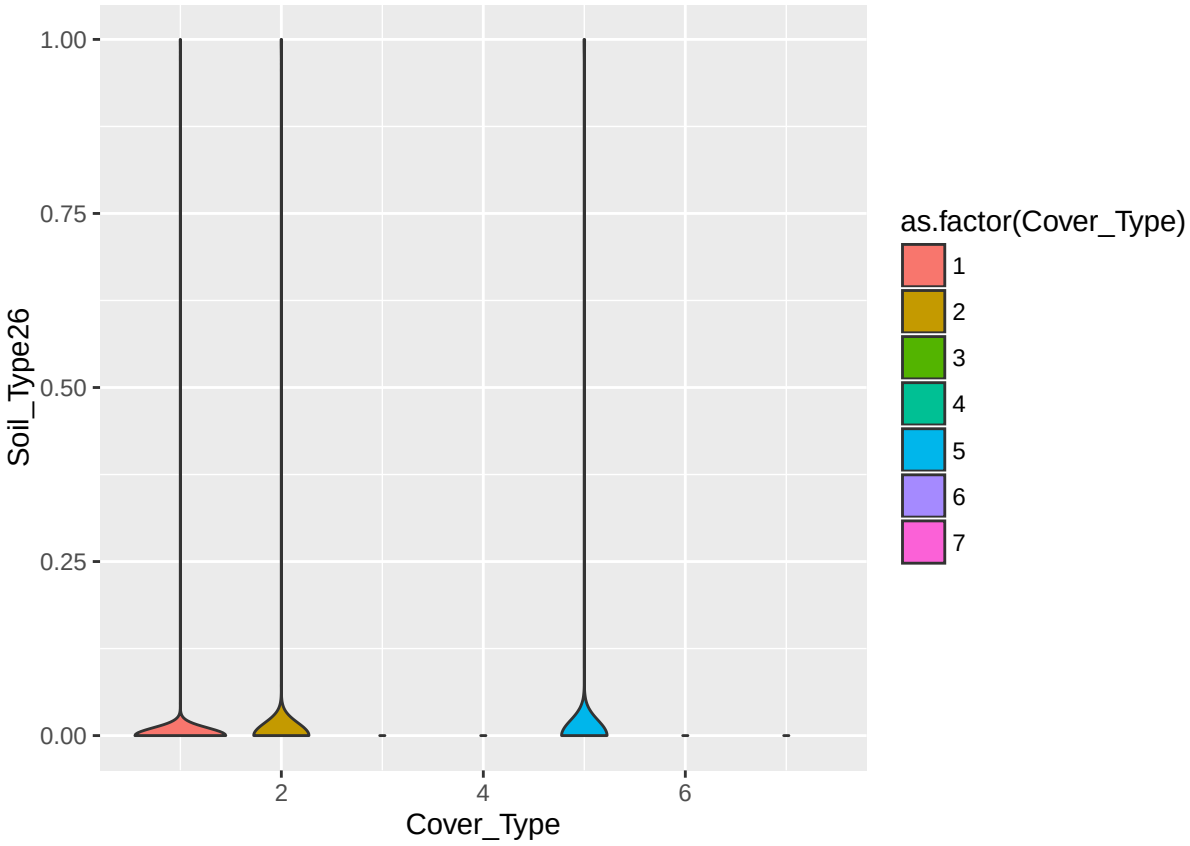

1.00
0.75
0.50
0.25
0.00
Soil_Type21
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

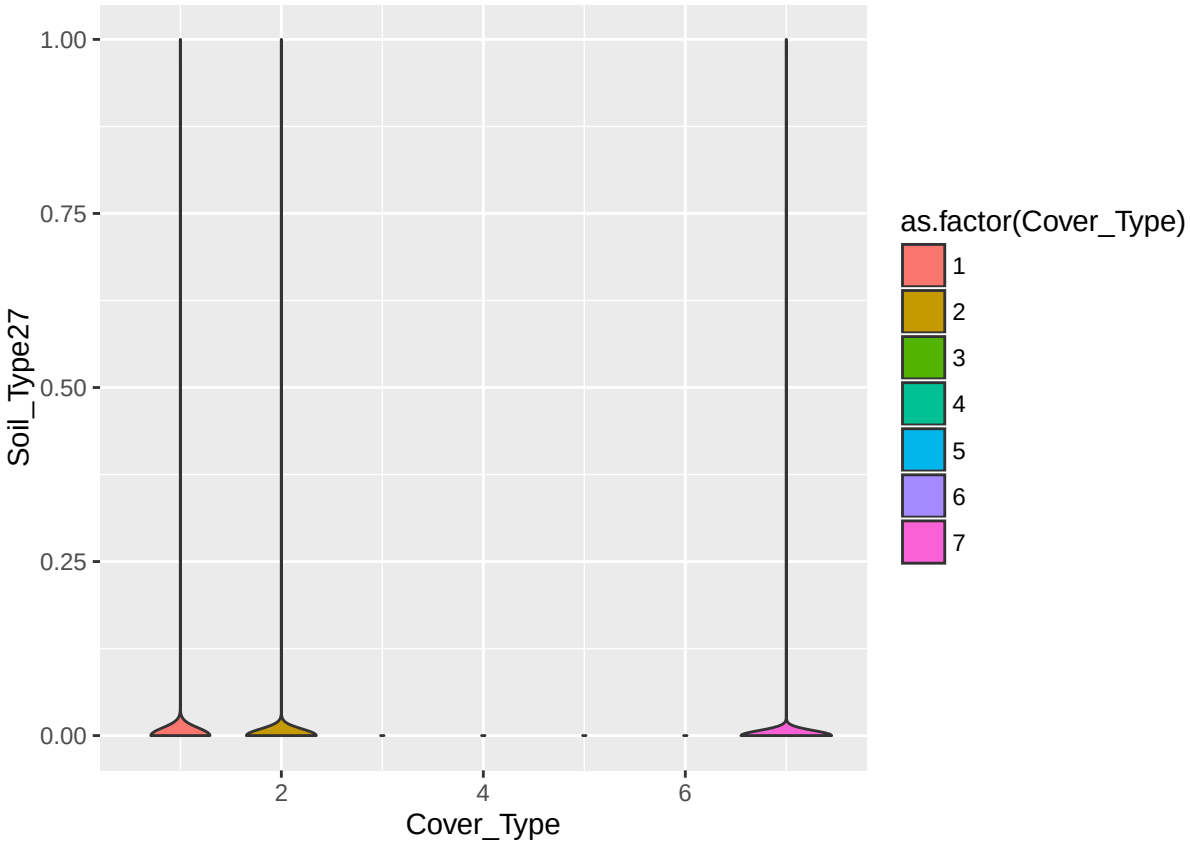

1.00
0.75
0.50
0.25
0.00
Soil_Type22
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

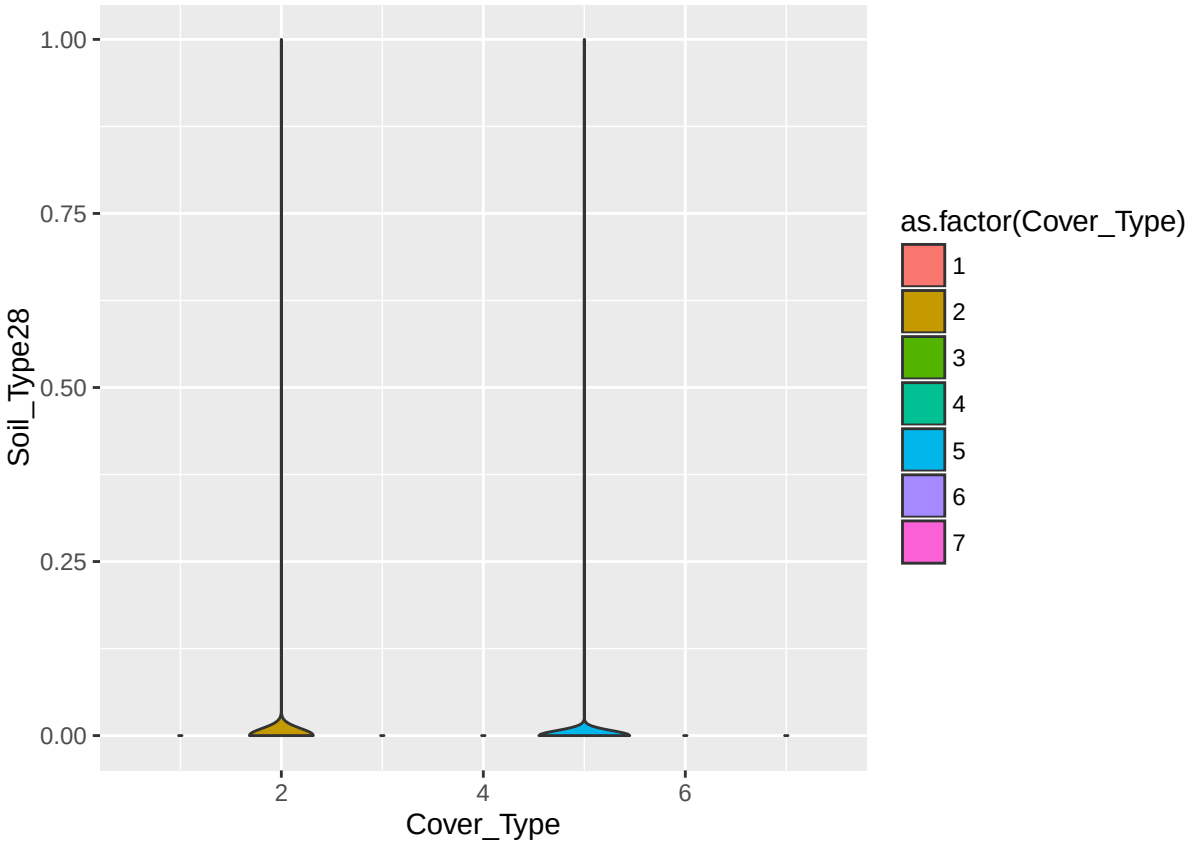

1.00
0.75
0.50
0.25
0.00
Soil_Type23
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

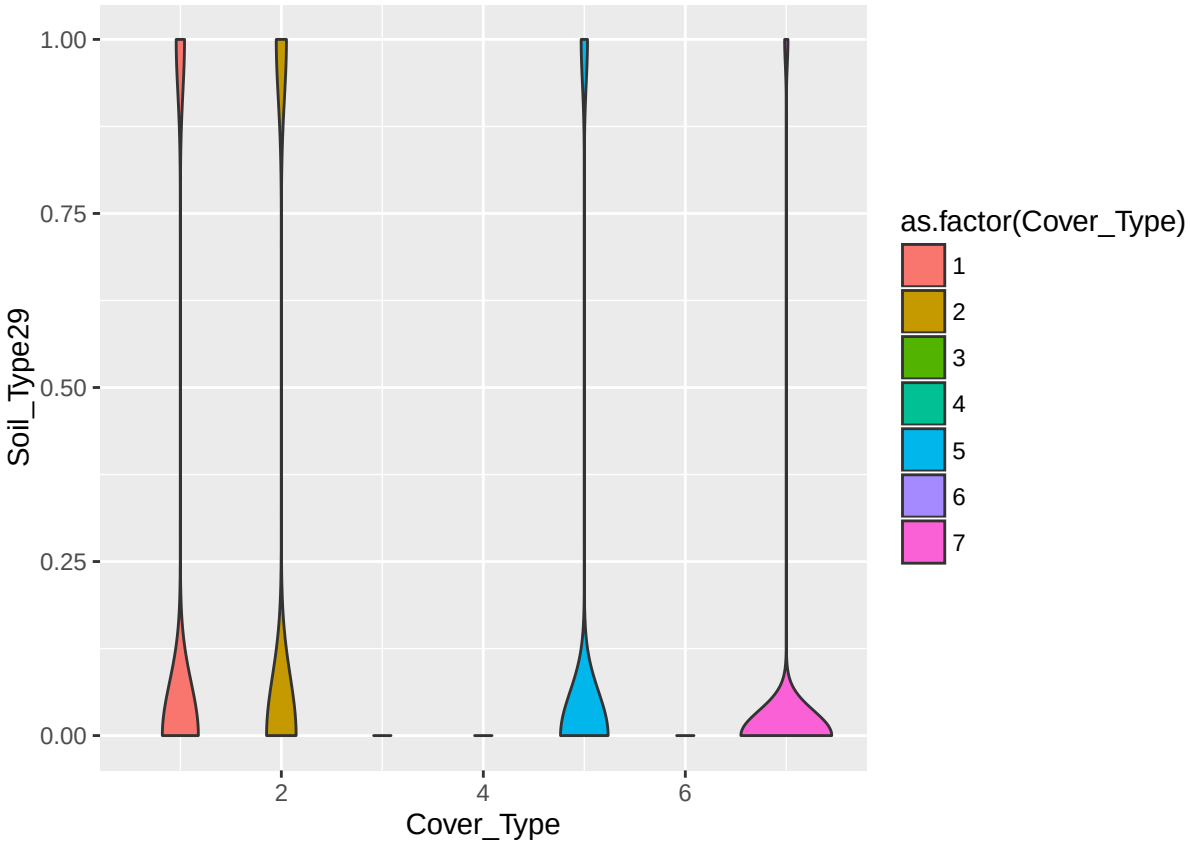

1.00
0.75
0.50
0.25
0.00
Soil_Type24
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

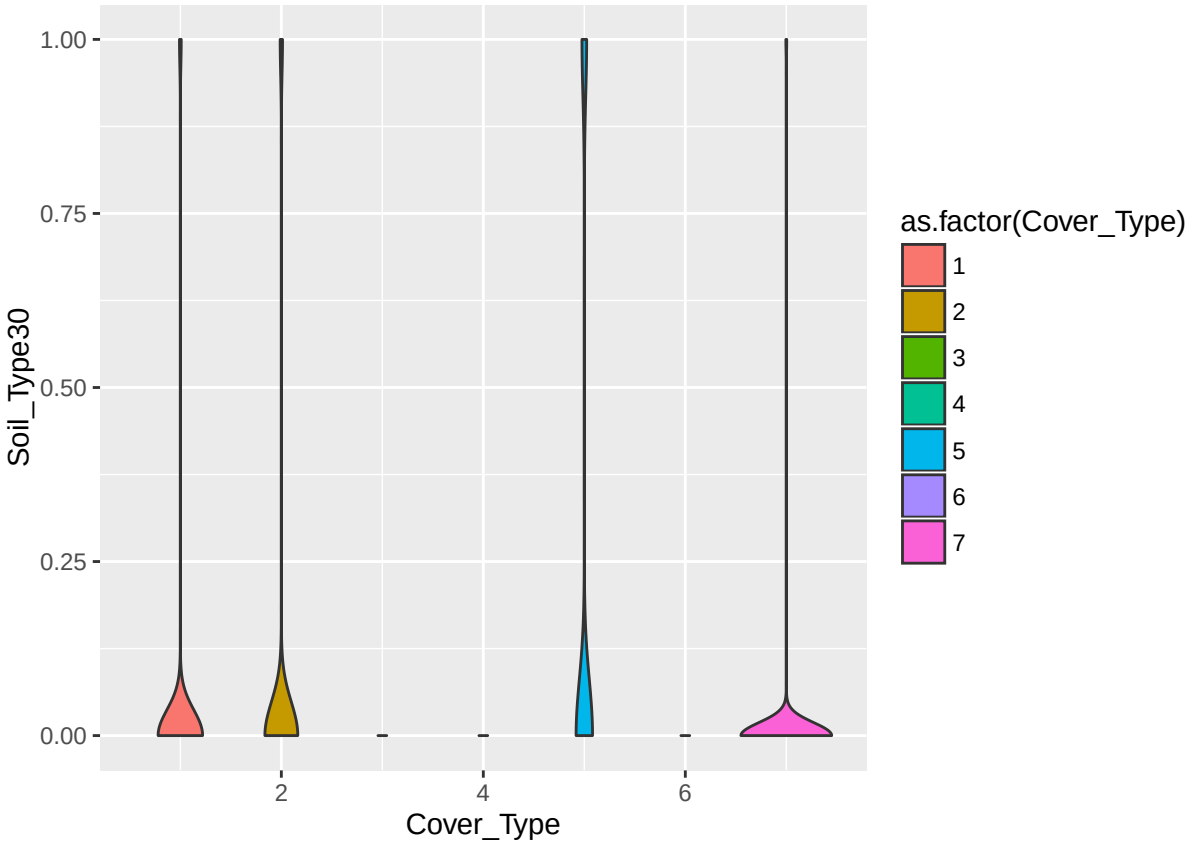

1.00
0.75
0.50
0.25
0.00
Soil_Type25
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

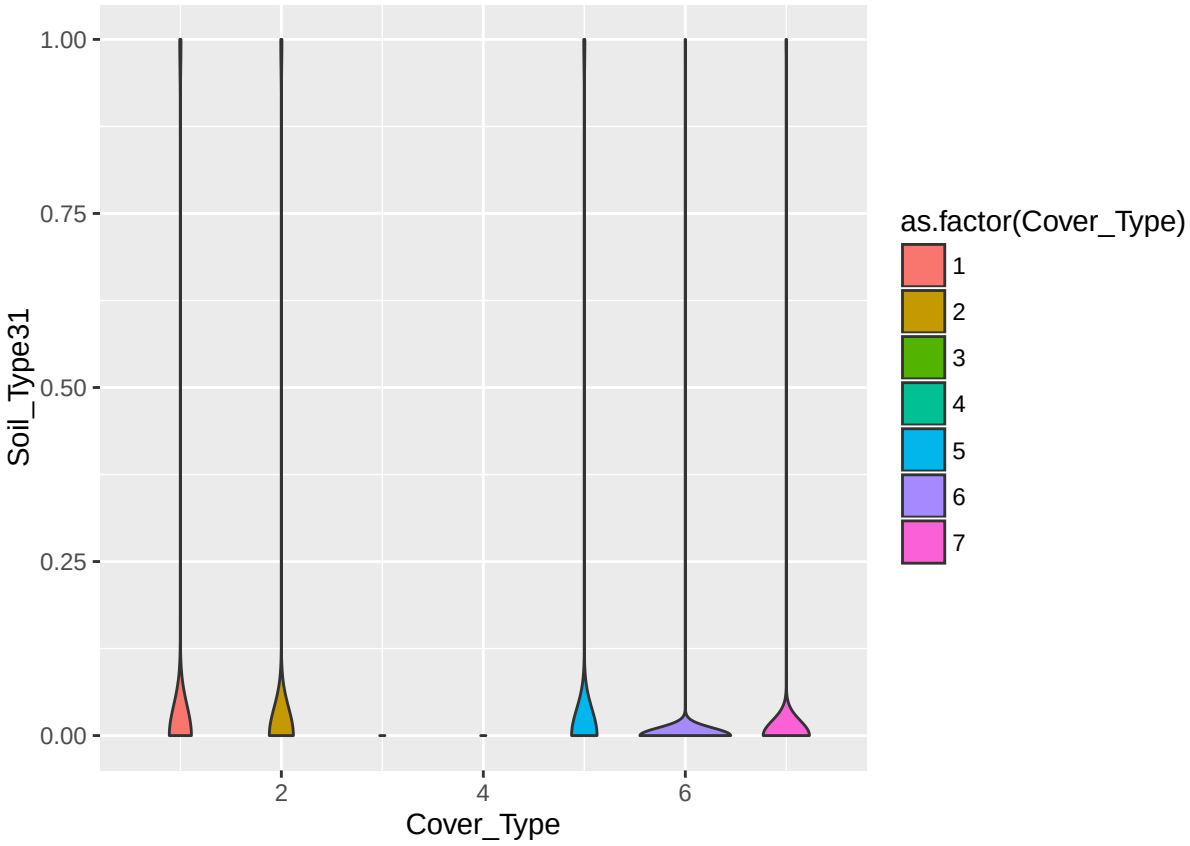

1.00
0.75
0.50
0.25
0.00
Soil_Type26
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

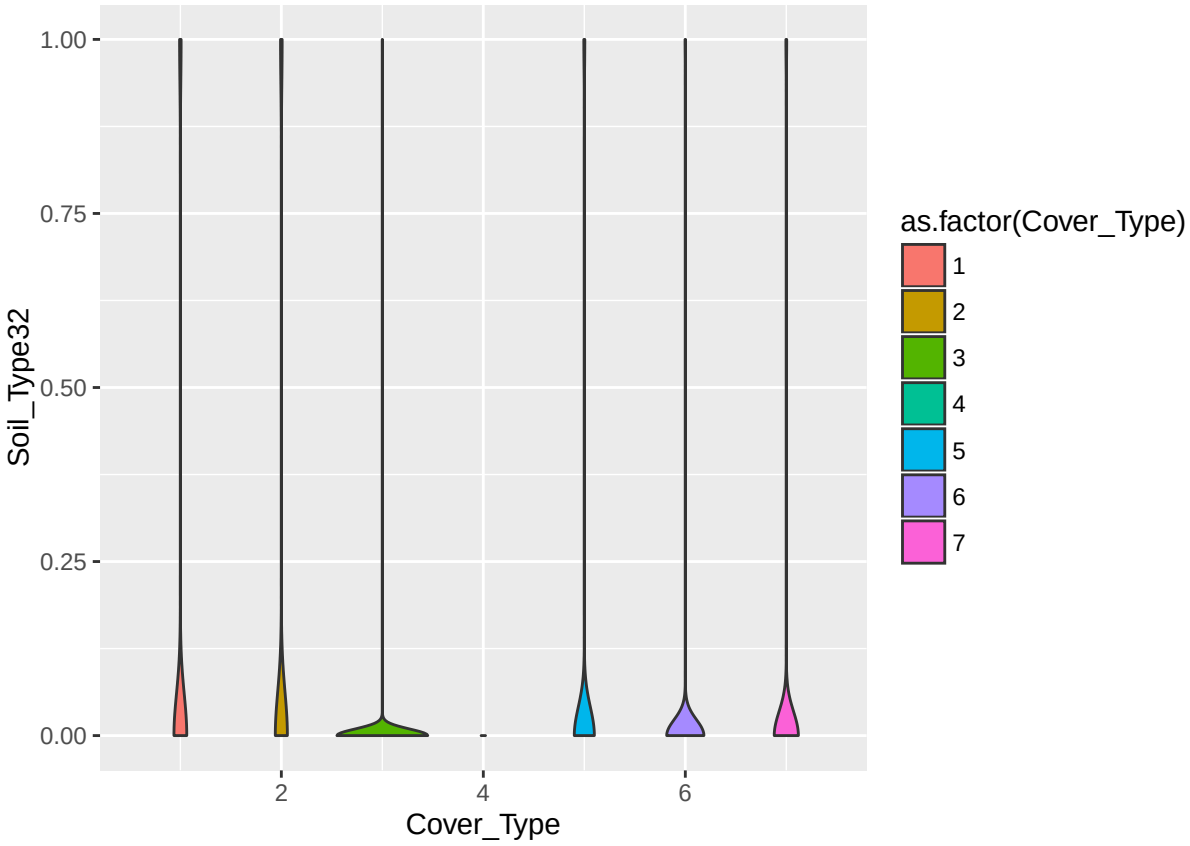

1.00
0.75
0.50
0.25
0.00
Soil_Type27
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

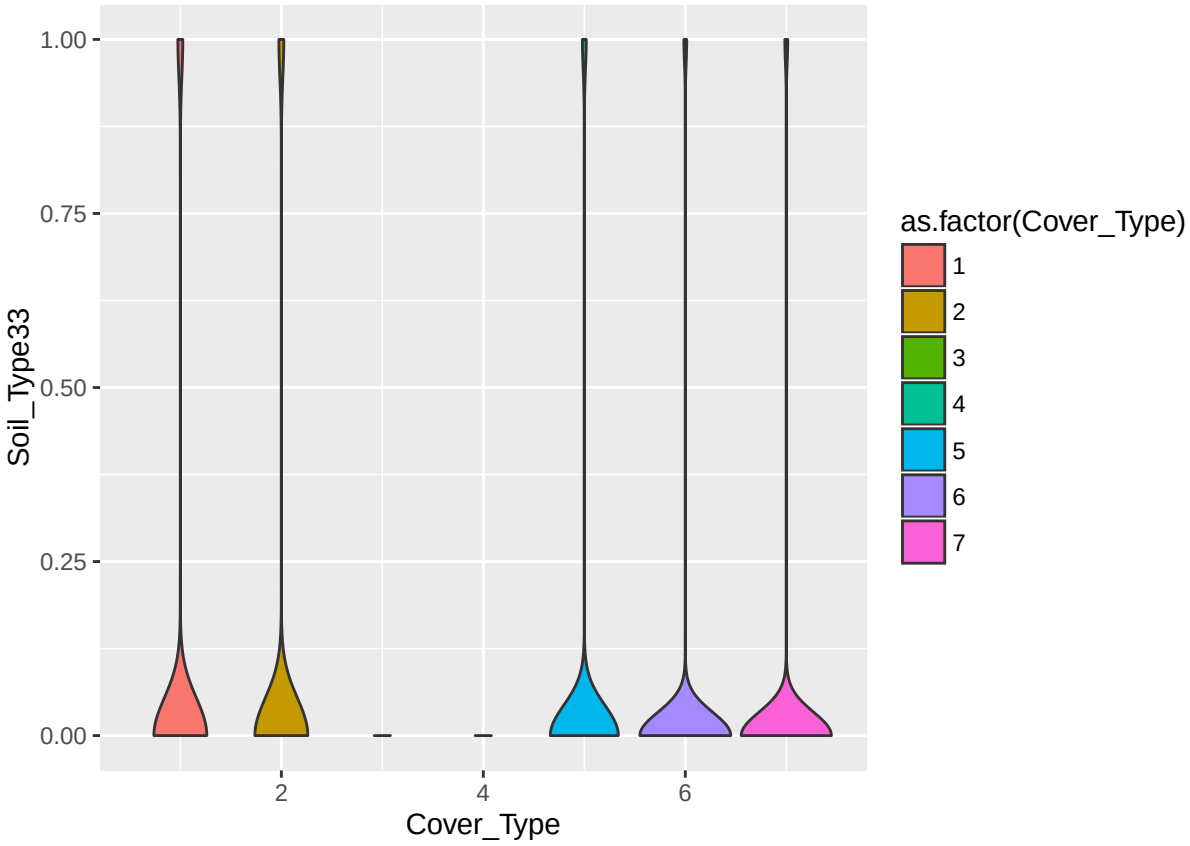

1.00
0.75
0.50
0.25
0.00
Soil_Type28
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

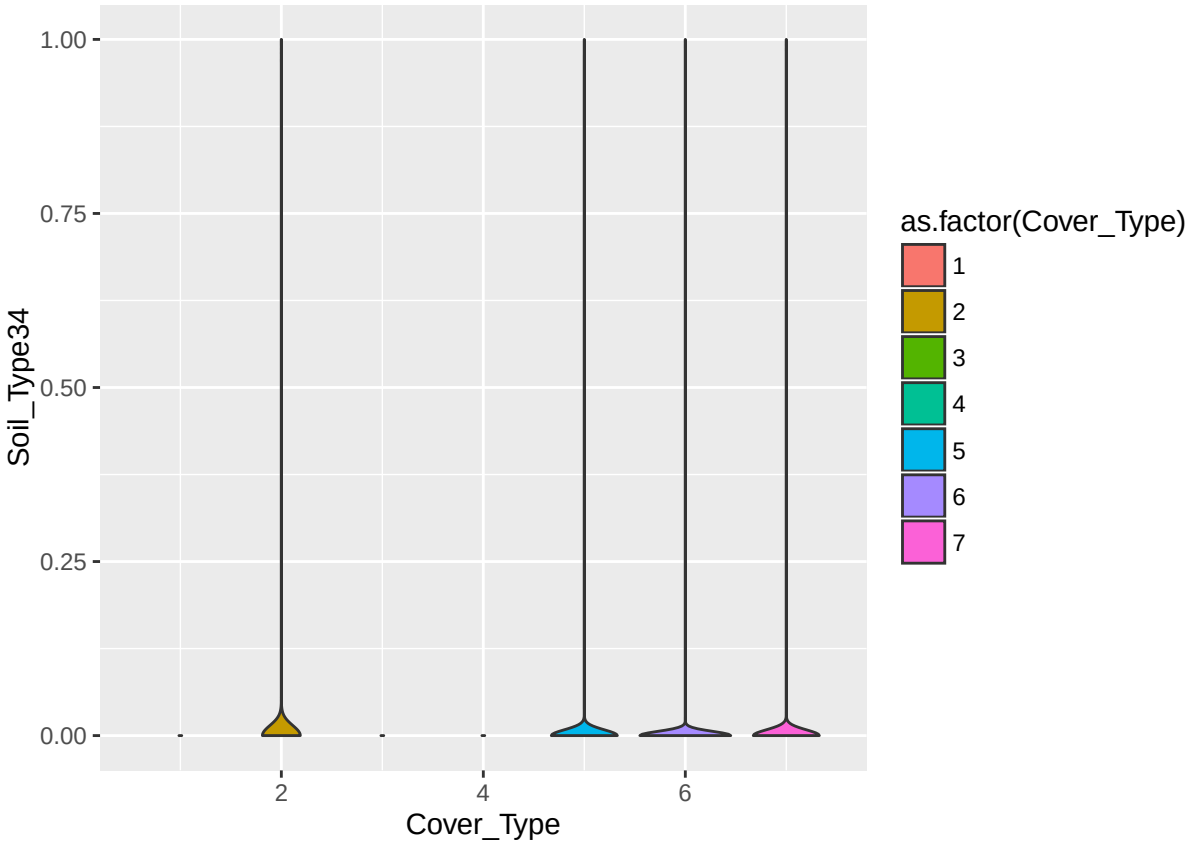

1.00
0.75
0.50
0.25
0.00
Soil_Type29
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

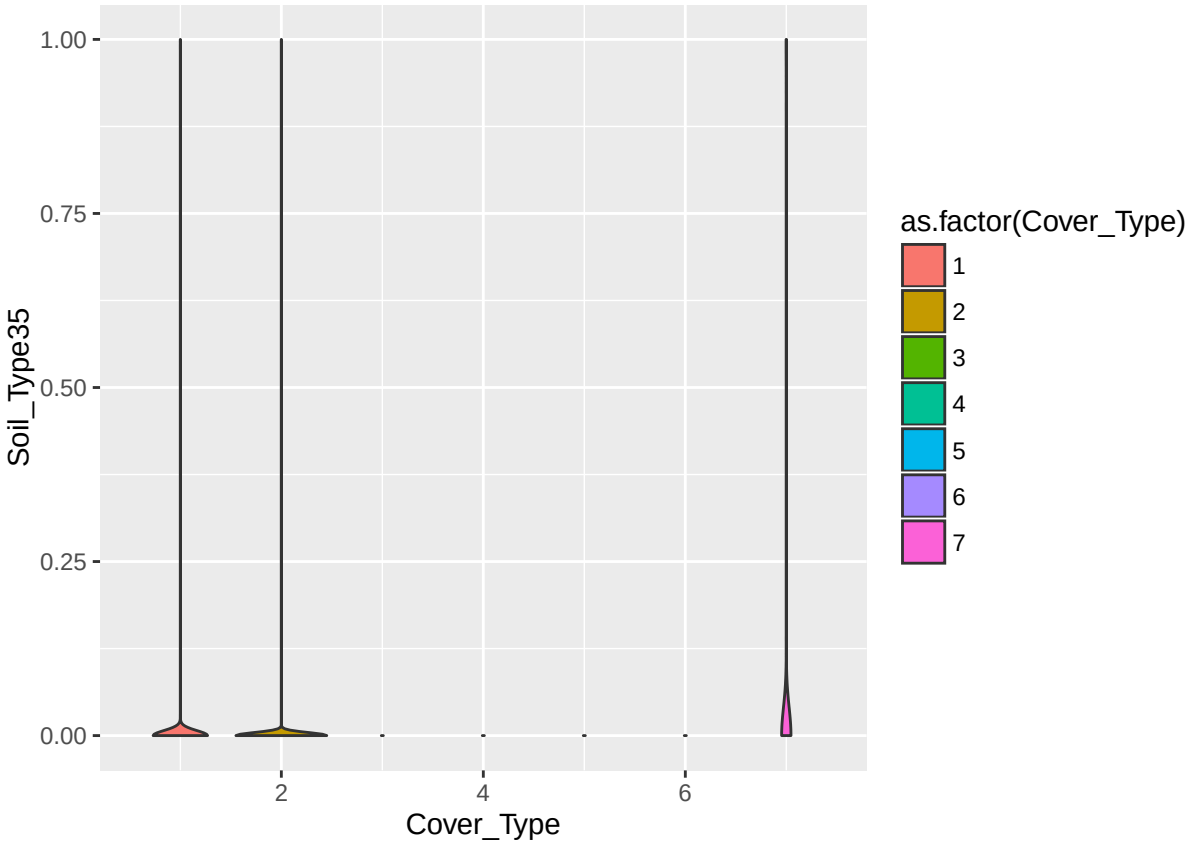

1.00
0.75
0.50
0.25
0.00
Soil_Type30
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

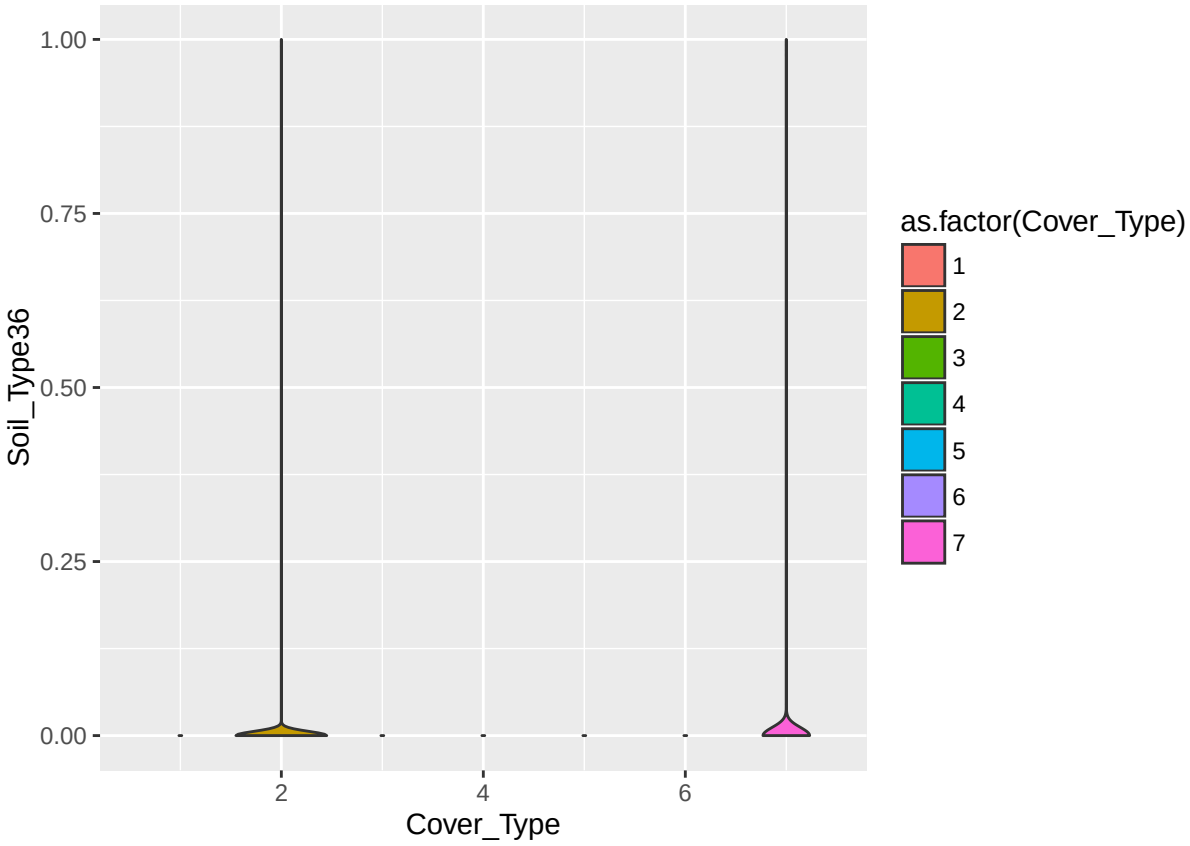

1.00
0.75
0.50
0.25
0.00
Soil_Type31
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

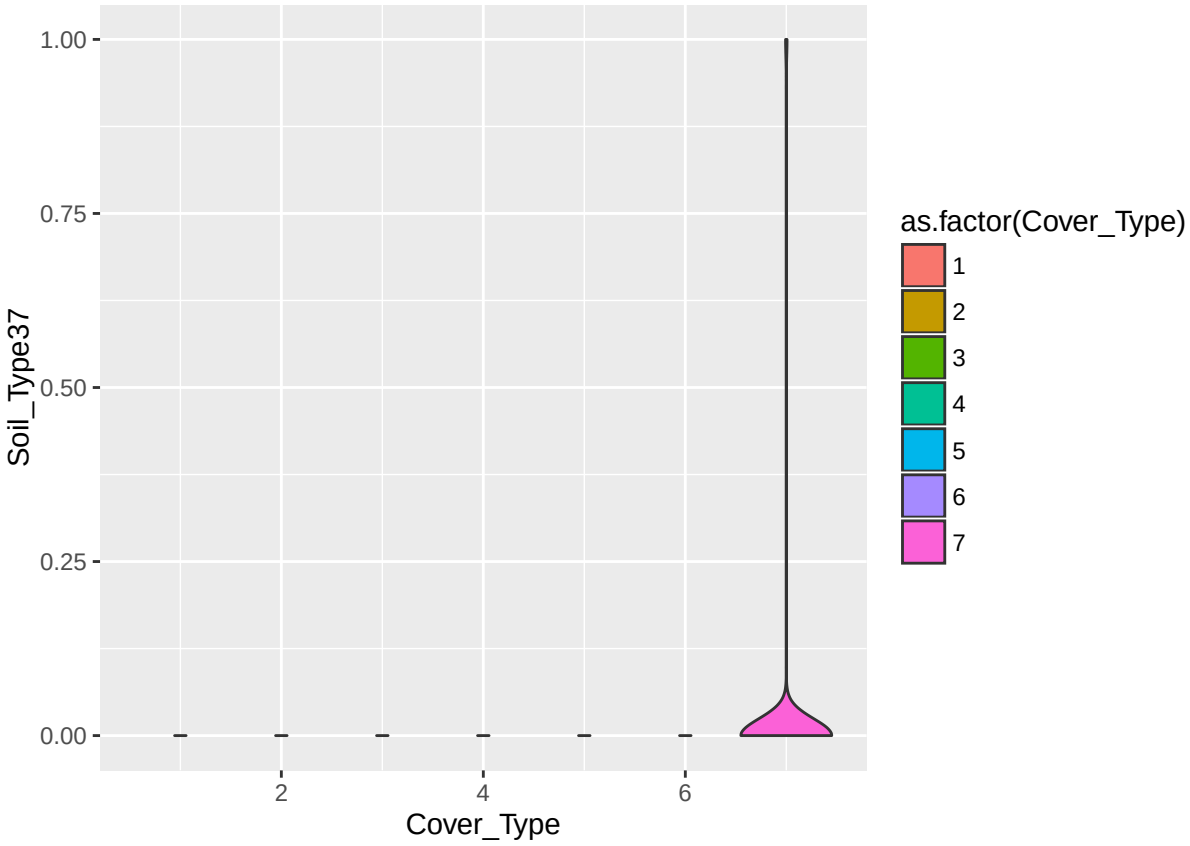

1.00
0.75
0.50
0.25
0.00
Soil_Type32
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

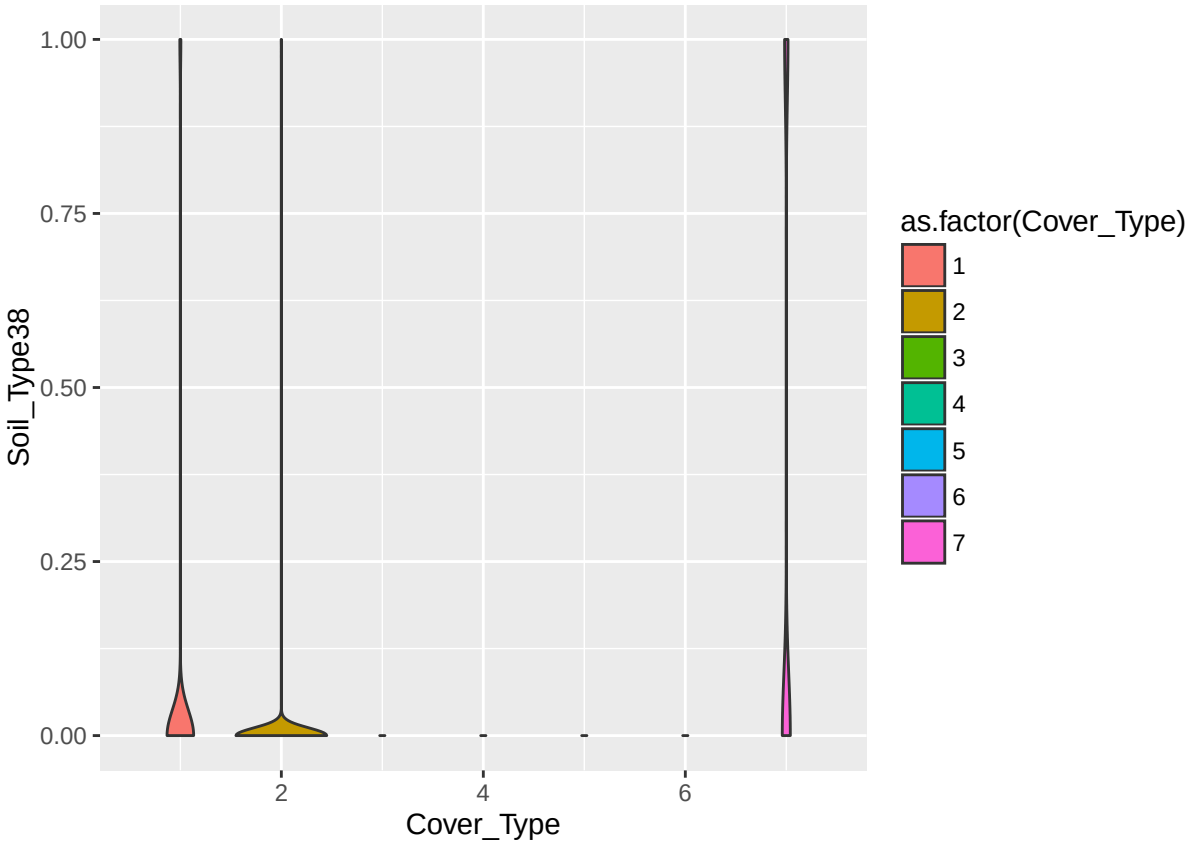

1.00
0.75
0.50
0.25
0.00
Soil_Type33
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

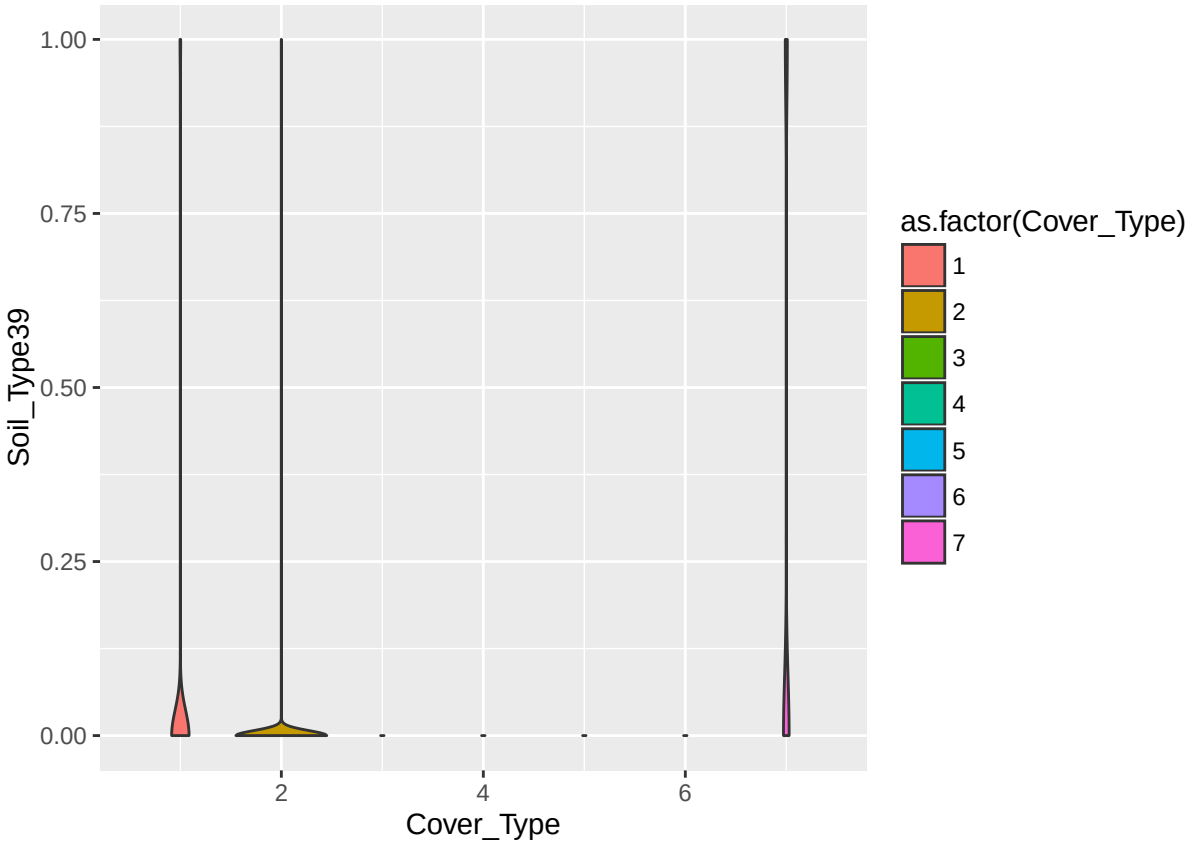

1.00
0.75
0.50
0.25
0.00
Soil_Type34
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

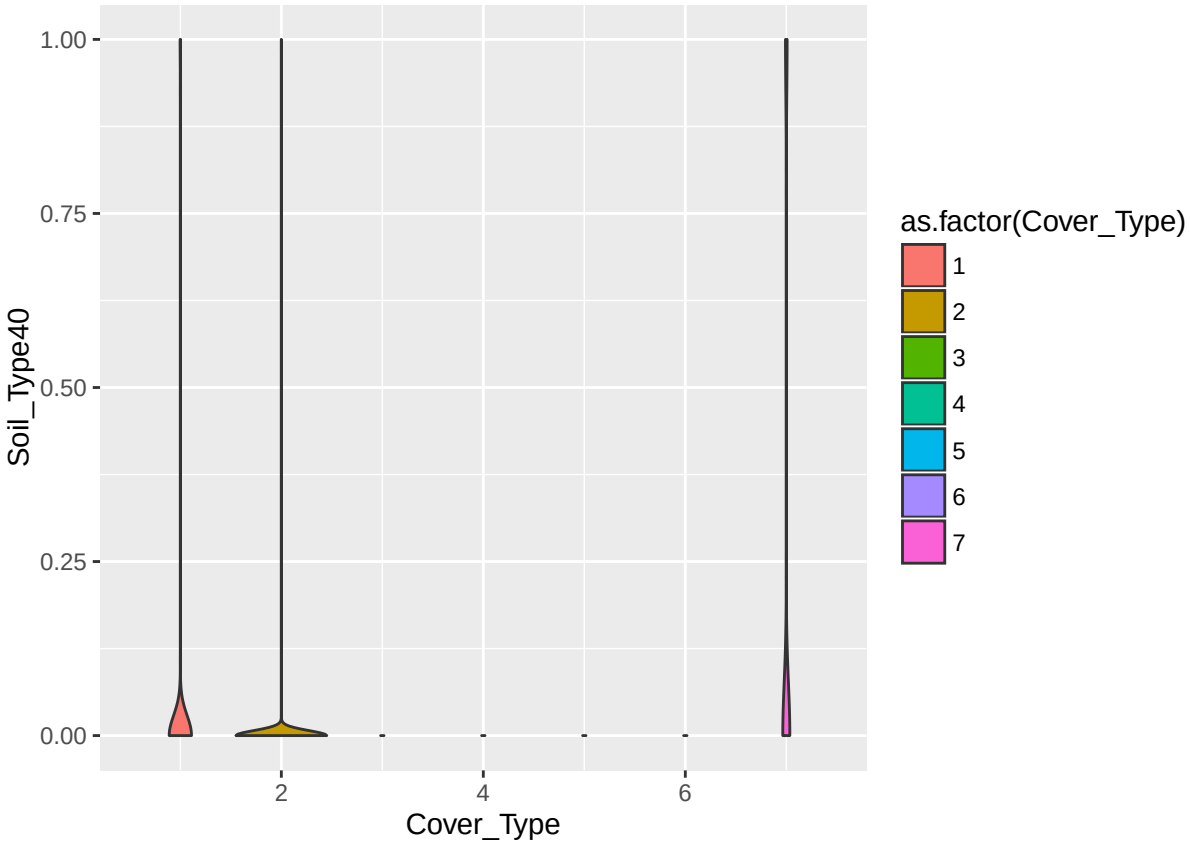

1.00
0.75
0.50
0.25
0.00
Soil_Type35
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

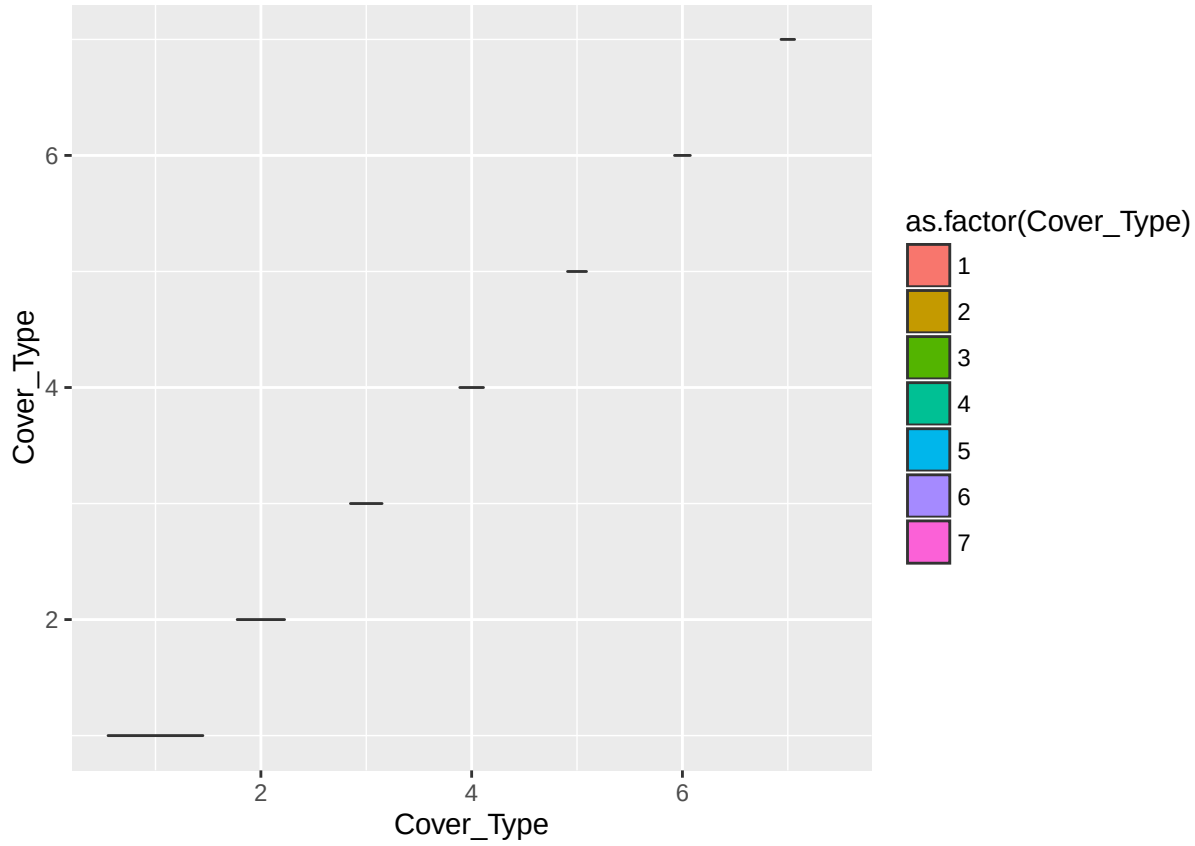

1.00
0.75
0.50
0.25
0.00
Soil_Type36
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7


1.00
0.75
0.50
0.25
0.00
Soil_Type37
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7

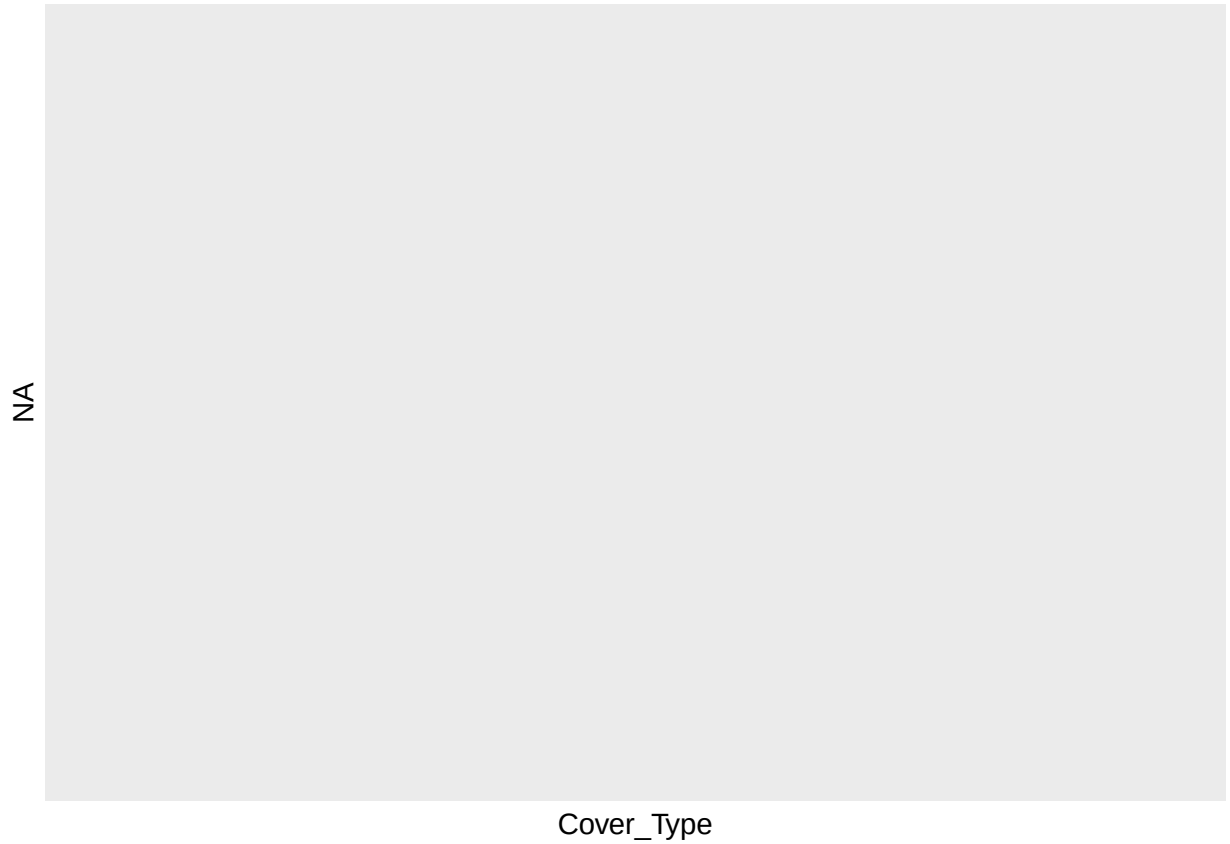

1.00
0.75
0.50
0.25
0.00
Soil_Type38
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7


1.00
0.75
0.50
0.25
0.00
Soil_Type39
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7


1.00
0.75
0.50
0.25
0.00
Soil_Type40
2
4
6
Cover_Type
as.factor(Cover_Type)
1
2
3
4
5
6
7




```
## Warning: Removed 15120 rows containing non-finite values (stat_ydensity).
```



###Prediction

```
which(lapply(train, sd)==0)
```

```
## named integer(0)
```

```
train <- as.data.frame(train)
train <- scale(train, center = TRUE)
head(train)
```

```
##        Elevation      Aspect      Slope Horizontal_Distance_To_Hydrology
## [1,] -0.36708298 -0.95994808 -1.5970788                        0.14663456
## [2,] -0.38144810 -0.91452896 -1.7153670                       -0.07233455
## [3,]  0.13090807 -0.16057160 -0.8873495                        0.19423654
## [4,]  0.08541851 -0.01523042  0.1772446                        0.07047139
## [5,] -0.36947717 -1.01445102 -1.7153670                       -0.35318623
## [6,] -0.40778417 -0.22415837 -1.2422141                        0.34656288
##      Vertical_Distance_To_Hydrology Horizontal_Distance_To_Roadways
## [1,]                     -0.8340466                       -0.9086513
## [2,]                     -0.9320228                       -0.9992128
## [3,]                      0.2273614                        1.1063422
## [4,]                      1.0928172                        1.0384210
## [5,]                     -0.8503760                       -0.9984581
## [6,]                     -1.0789870                       -1.2429741
##      Hillshade_9am Hillshade_Noon Hillshade_3pm
## [1,]     0.2714448      0.5716346    0.28124958
```

```
## [2,]     0.2387236      0.7032022    0.34661591
## [3,]     0.6968195      0.8347698   -0.00200451
## [4,]     0.8277041      0.8347698   -0.28525860
## [5,]     0.2387236      0.6593463    0.32482713
## [6,]     0.5659350      0.7909139    0.10693937
##      Horizontal_Distance_To_Fire_Points Wilderness_Area1 Wilderness_Area2
## [1,]                           4.334662         1.789774       -0.1847342
## [2,]                           4.285568         1.789774       -0.1847342
## [3,]                           4.191017         1.789774       -0.1847342
## [4,]                           4.272840         1.789774       -0.1847342
## [5,]                           4.237383         1.789774       -0.1847342
## [6,]                           4.109194         1.789774       -0.1847342
##      Wilderness_Area3 Wilderness_Area4 Soil_Type1 Soil_Type2 Soil_Type3
## [1,]       -0.8507731        -0.668994  -0.155054 -0.2072958 -0.2606587
## [2,]       -0.8507731        -0.668994  -0.155054 -0.2072958 -0.2606587
## [3,]       -0.8507731        -0.668994  -0.155054 -0.2072958 -0.2606587
## [4,]       -0.8507731        -0.668994  -0.155054 -0.2072958 -0.2606587
## [5,]       -0.8507731        -0.668994  -0.155054 -0.2072958 -0.2606587
## [6,]       -0.8507731        -0.668994  -0.155054 -0.2072958 -0.2606587
##      Soil_Type4 Soil_Type5 Soil_Type6   Soil_Type8  Soil_Type9 Soil_Type10
## [1,] -0.2429858 -0.1050351 -0.2119376 -0.008132501 -0.02572488  -0.4062482
## [2,] -0.2429858 -0.1050351 -0.2119376 -0.008132501 -0.02572488  -0.4062482
## [3,] -0.2429858 -0.1050351 -0.2119376 -0.008132501 -0.02572488  -0.4062482
## [4,] -0.2429858 -0.1050351 -0.2119376 -0.008132501 -0.02572488  -0.4062482
## [5,] -0.2429858 -0.1050351 -0.2119376 -0.008132501 -0.02572488  -0.4062482
## [6,] -0.2429858 -0.1050351 -0.2119376 -0.008132501 -0.02572488  -0.4062482
##      Soil_Type11 Soil_Type12 Soil_Type13 Soil_Type14 Soil_Type16
## [1,]  -0.1661052  -0.1234547  -0.1802849  -0.1063148 -0.08715767
## [2,]  -0.1661052  -0.1234547  -0.1802849  -0.1063148 -0.08715767
## [3,]  -0.1661052   8.0996043  -0.1802849  -0.1063148 -0.08715767
## [4,]  -0.1661052  -0.1234547  -0.1802849  -0.1063148 -0.08715767
## [5,]  -0.1661052  -0.1234547  -0.1802849  -0.1063148 -0.08715767
## [6,]  -0.1661052  -0.1234547  -0.1802849  -0.1063148 -0.08715767
##      Soil_Type17 Soil_Type18 Soil_Type19 Soil_Type20 Soil_Type21
## [1,]  -0.2053797 -0.06311735 -0.05523957 -0.09632137 -0.03254615
## [2,]  -0.2053797 -0.06311735 -0.05523957 -0.09632137 -0.03254615
## [3,]  -0.2053797 -0.06311735 -0.05523957 -0.09632137 -0.03254615
## [4,]  -0.2053797 -0.06311735 -0.05523957 -0.09632137 -0.03254615
## [5,]  -0.2053797 -0.06311735 -0.05523957 -0.09632137 -0.03254615
## [6,]  -0.2053797 -0.06311735 -0.05523957 -0.09632137 -0.03254615
##      Soil_Type22 Soil_Type23 Soil_Type24  Soil_Type25 Soil_Type26
## [1,]  -0.1528028  -0.2295678  -0.1314919 -0.008132501 -0.05986645
## [2,]  -0.1528028  -0.2295678  -0.1314919 -0.008132501 -0.05986645
## [3,]  -0.1528028  -0.2295678  -0.1314919 -0.008132501 -0.05986645
## [4,]  -0.1528028  -0.2295678  -0.1314919 -0.008132501 -0.05986645
## [5,]  -0.1528028  -0.2295678  -0.1314919 -0.008132501 -0.05986645
## [6,]  -0.1528028  -0.2295678  -0.1314919 -0.008132501 -0.05986645
##      Soil_Type27 Soil_Type28 Soil_Type29 Soil_Type30 Soil_Type31
## [1,] -0.03151163 -0.02440396   3.2727882  -0.2244134  -0.1498304
## [2,] -0.03151163 -0.02440396   3.2727882  -0.2244134  -0.1498304
## [3,] -0.03151163 -0.02440396  -0.3055297  -0.2244134  -0.1498304
## [4,] -0.03151163 -0.02440396  -0.3055297   4.4557670  -0.1498304
## [5,] -0.03151163 -0.02440396   3.2727882  -0.2244134  -0.1498304
## [6,] -0.03151163 -0.02440396   3.2727882  -0.2244134  -0.1498304
```

```
##      Soil_Type32 Soil_Type33 Soil_Type34 Soil_Type35 Soil_Type36
## [1,]  -0.2186639  -0.2060782 -0.03817133 -0.08240995 -0.02572488
## [2,]  -0.2186639  -0.2060782 -0.03817133 -0.08240995 -0.02572488
## [3,]  -0.2186639  -0.2060782 -0.03817133 -0.08240995 -0.02572488
## [4,]  -0.2186639  -0.2060782 -0.03817133 -0.08240995 -0.02572488
## [5,]  -0.2186639  -0.2060782 -0.03817133 -0.08240995 -0.02572488
## [6,]  -0.2186639  -0.2060782 -0.03817133 -0.08240995 -0.02572488
##      Soil_Type37 Soil_Type38 Soil_Type39 Soil_Type40 Cover_Type
## [1,] -0.04747206  -0.2249007  -0.2131273  -0.1769335  0.4999835
## [2,] -0.04747206  -0.2249007  -0.2131273  -0.1769335  0.4999835
## [3,] -0.04747206  -0.2249007  -0.2131273  -0.1769335 -0.9999669
## [4,] -0.04747206  -0.2249007  -0.2131273  -0.1769335 -0.9999669
## [5,] -0.04747206  -0.2249007  -0.2131273  -0.1769335  0.4999835
## [6,] -0.04747206  -0.2249007  -0.2131273  -0.1769335 -0.9999669
```

```
test <- fread("test.csv", header = TRUE)
```

```
##
Read 84.8% of 565892 rows
Read 565892 rows and 55 (of 55) columns from 0.071 GB file in 00:00:03
```

```
test <- test[, Id:=NULL]

require(randomForest)
```

```
## Loading required package: randomForest

## Warning: package 'randomForest' was built under R version 3.3.1

## randomForest 4.6-12

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':
##
##     margin
```

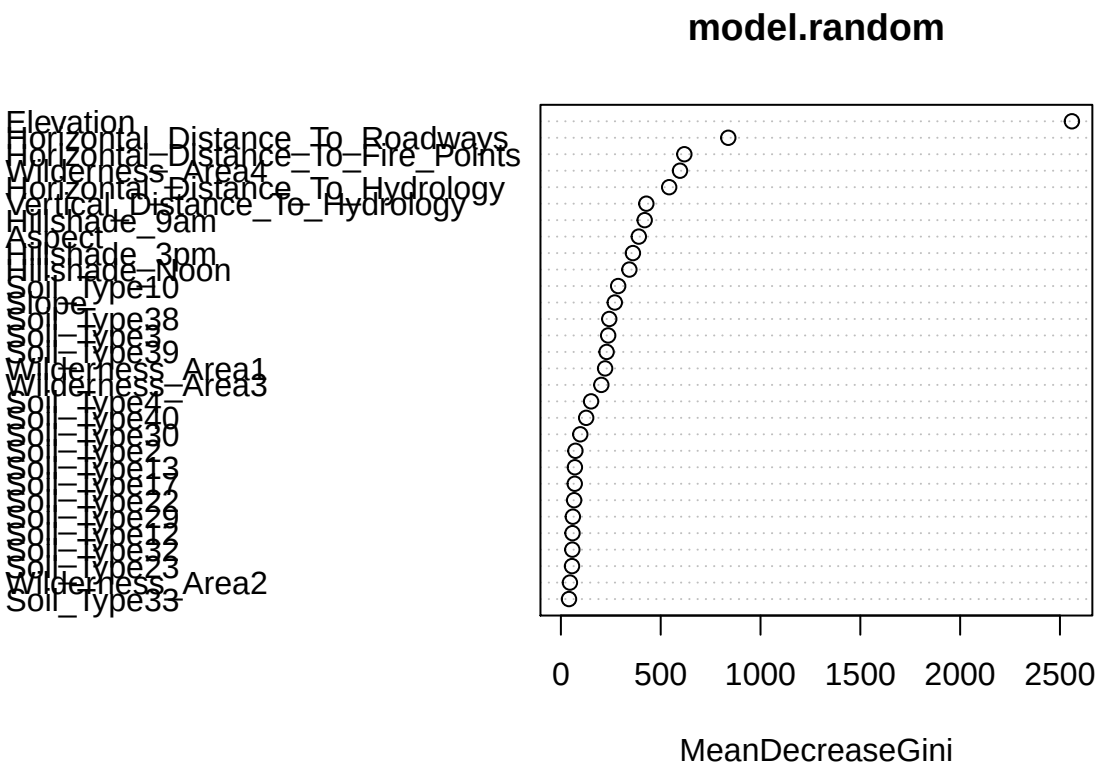
```
class(train)
```

```
## [1] "matrix"
```

```
names(train)
```

```
## NULL
```

```
train <- as.data.table(train)
model.random <- randomForest(as.factor(Cover_Type)~., train)
varImpPlot(model.random)
```




```
test$Cover_Type <- predict(model.random, test, class= "type")

head(test$Cover_Type)
```

```
## [1] -1.49995039600521 -1.49995039600521 -1.49995039600521 -1.49995039600521
## [5] -1.49995039600521 -1.49995039600521
## 7 Levels: -1.49995039600521 -0.999966930670141 -0.49998346533507 ... 1.49995039600521
```