Linear regression on trees

```
data_path = "ble.csv"
mydata = read.table(file = data_path, sep=";", dec=".", header= TRUE)
mydata
```

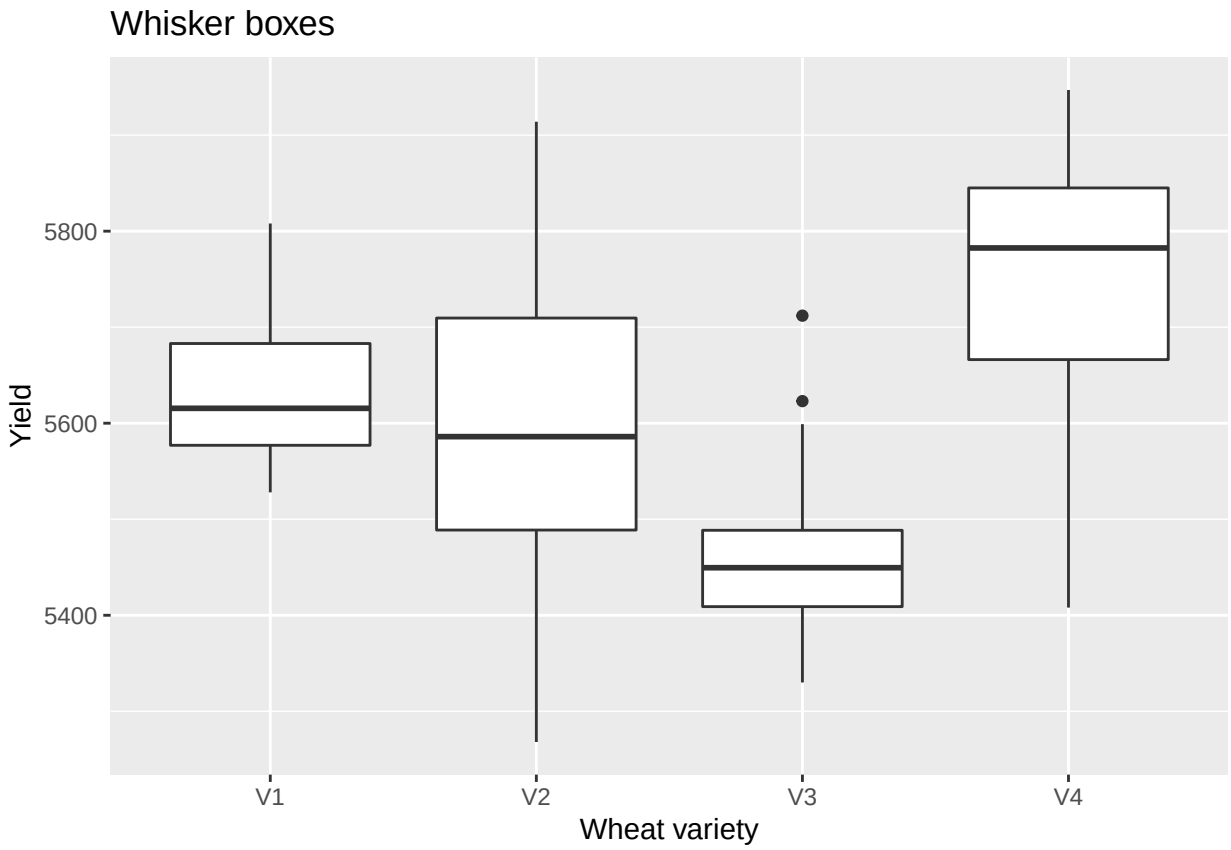
```
##    parcelle variete phyto  rdt
## 1         1      V1  Avec 5652
## 2         2      V1  Avec 5583
## 3         3      V1  Avec 5612
## 4         4      V1  Avec 5735
## 5         5      V1  Avec 5704
## 6         6      V1  Avec 5544
## 7         7      V1  Avec 5563
## 8         8      V1  Avec 5610
## 9         9      V1  Avec 5641
## 10       10      V1  Avec 5637
## 11       11      V1  Sans 5581
## 12       12      V1  Sans 5808
## 13       13      V1  Sans 5582
## 14       14      V1  Sans 5528
## 15       15      V1  Sans 5754
## 16       16      V1  Sans 5676
## 17       17      V1  Sans 5558
## 18       18      V1  Sans 5724
## 19       19      V1  Sans 5619
## 20       20      V1  Sans 5565
## 21       21      V2  Avec 5458
## 22       22      V2  Avec 5591
## 23       23      V2  Avec 5501
## 24       24      V2  Avec 5714
## 25       25      V2  Avec 5708
## 26       26      V2  Avec 5731
## 27       27      V2  Avec 5691
## 28       28      V2  Avec 5571
## 29       29      V2  Avec 5613
## 30       30      V2  Avec 5359
## 31       31      V2  Sans 5744
## 32       32      V2  Sans 5771
## 33       33      V2  Sans 5592
## 34       34      V2  Sans 5499
## 35       35      V2  Sans 5437
## 36       36      V2  Sans 5413
## 37       37      V2  Sans 5581
## 38       38      V2  Sans 5268
## 39       39      V2  Sans 5526
## 40       40      V2  Sans 5914
## 41       41      V3  Avec 5330
## 42       42      V3  Avec 5485
## 43       43      V3  Avec 5712
## 44       44      V3  Avec 5499
## 45       45      V3  Avec 5461
## 46       46      V3  Avec 5444
```

```
## 47       47      V3  Avec 5412
## 48       48      V3  Avec 5396
## 49       49      V3  Avec 5466
## 50       50      V3  Avec 5400
## 51       51      V3  Sans 5431
## 52       52      V3  Sans 5599
## 53       53      V3  Sans 5455
## 54       54      V3  Sans 5435
## 55       55      V3  Sans 5371
## 56       56      V3  Sans 5499
## 57       57      V3  Sans 5623
## 58       58      V3  Sans 5367
## 59       59      V3  Sans 5432
## 60       60      V3  Sans 5475
## 61       61      V4  Avec 5844
## 62       62      V4  Avec 5713
## 63       63      V4  Avec 5841
## 64       64      V4  Avec 5716
## 65       65      V4  Avec 5803
## 66       66      V4  Avec 5725
## 67       67      V4  Avec 5881
## 68       68      V4  Avec 5672
## 69       69      V4  Avec 5869
## 70       70      V4  Avec 5602
## 71       71      V4  Sans 5766
## 72       72      V4  Sans 5408
## 73       73      V4  Sans 5947
## 74       74      V4  Sans 5644
## 75       75      V4  Sans 5848
## 76       76      V4  Sans 5809
## 77       77      V4  Sans 5627
## 78       78      V4  Sans 5881
## 79       79      V4  Sans 5649
## 80       80      V4  Sans 5799
```

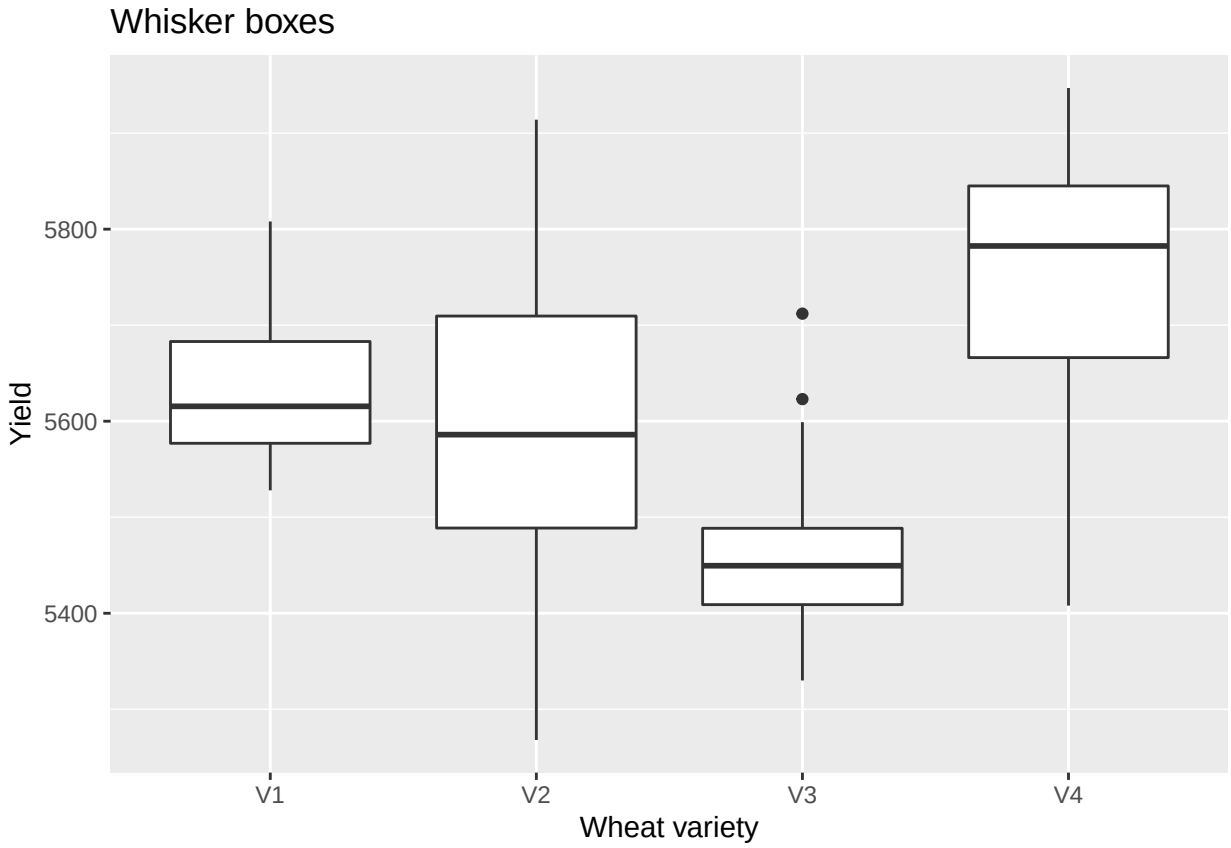
1- factor ANOVA

```
library(ggplot2)

ggplot(mydata, aes(x=variete,y=rdt)) +
  geom_boxplot() +
  ggtitle("Whisker boxes") +
  xlab("Wheat variety") +
  ylab("Yield")
```




```
ggplot(mydata, aes(x=variete,y=rdt)) +
  geom_boxplot() +
  ggtitle("Whisker boxes") +
  xlab("Wheat variety") +
  ylab("Yield")
```




```
annova_variete <- lm(rdt~variete, data=mydata)
summary(annova_variete)
```

```
##
## Call:
## lm(formula = rdt ~ variete, data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -344.20  -69.30   -6.60   89.15  329.90
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5633.80      26.30 214.211  < 2e-16 ***
## varieteV2     -49.70      37.19  -1.336  0.18546
## varieteV3    -169.20      37.19  -4.549    2e-05 ***
## varieteV4     118.40      37.19   3.183  0.00211 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 117.6 on 76 degrees of freedom
## Multiple R-squared:  0.4476, Adjusted R-squared:  0.4258
## F-statistic: 20.53 on 3 and 76 DF,  p-value: 7.674e-10
```

```
anova(annova_variete)
```

```
## Analysis of Variance Table
##
```

```
## Response: rdt
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## variete    3  851845  283948  20.525 7.674e-10 ***
## Residuals 76 1051387   13834
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
annova_variete_phyto <- lm(rdt~variete*phyto, data=mydata)
summary(annova_variete_phyto)
```

```
##
## Call:
## lm(formula = rdt ~ variete * phyto, data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -329.80  -67.45   -8.20   76.28  339.50
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          5628.10      38.09 147.772  < 2e-16 ***
## varieteV2             -34.40      53.86  -0.639  0.52507
## varieteV3            -167.60      53.86  -3.112  0.00267 **
## varieteV4             138.50      53.86   2.571  0.01219 *
## phytoSans              11.40      53.86   0.212  0.83298
## varieteV2:phytoSans   -30.60      76.17  -0.402  0.68908
## varieteV3:phytoSans    -3.20      76.17  -0.042  0.96661
## varieteV4:phytoSans   -40.20      76.17  -0.528  0.59930
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 120.4 on 72 degrees of freedom
## Multiple R-squared:  0.4512, Adjusted R-squared:  0.3979
## F-statistic: 8.458 on 7 and 72 DF,  p-value: 1.622e-07
```

```
anova(annova_variete_phyto)
```

```
## Analysis of Variance Table
##
## Response: rdt
##               Df  Sum Sq Mean Sq F value    Pr(>F)
## variete        3  851845  283948 19.5749 2.205e-09 ***
## phyto          1    1008    1008  0.0695    0.7928
## variete:phyto  3    5968    1989  0.1371    0.9375
## Residuals     72 1044411   14506
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can keep the value depending of the variety, the others no effects So the model is mainly drived by the variety