```
library(ggplot2)
```

Loading data

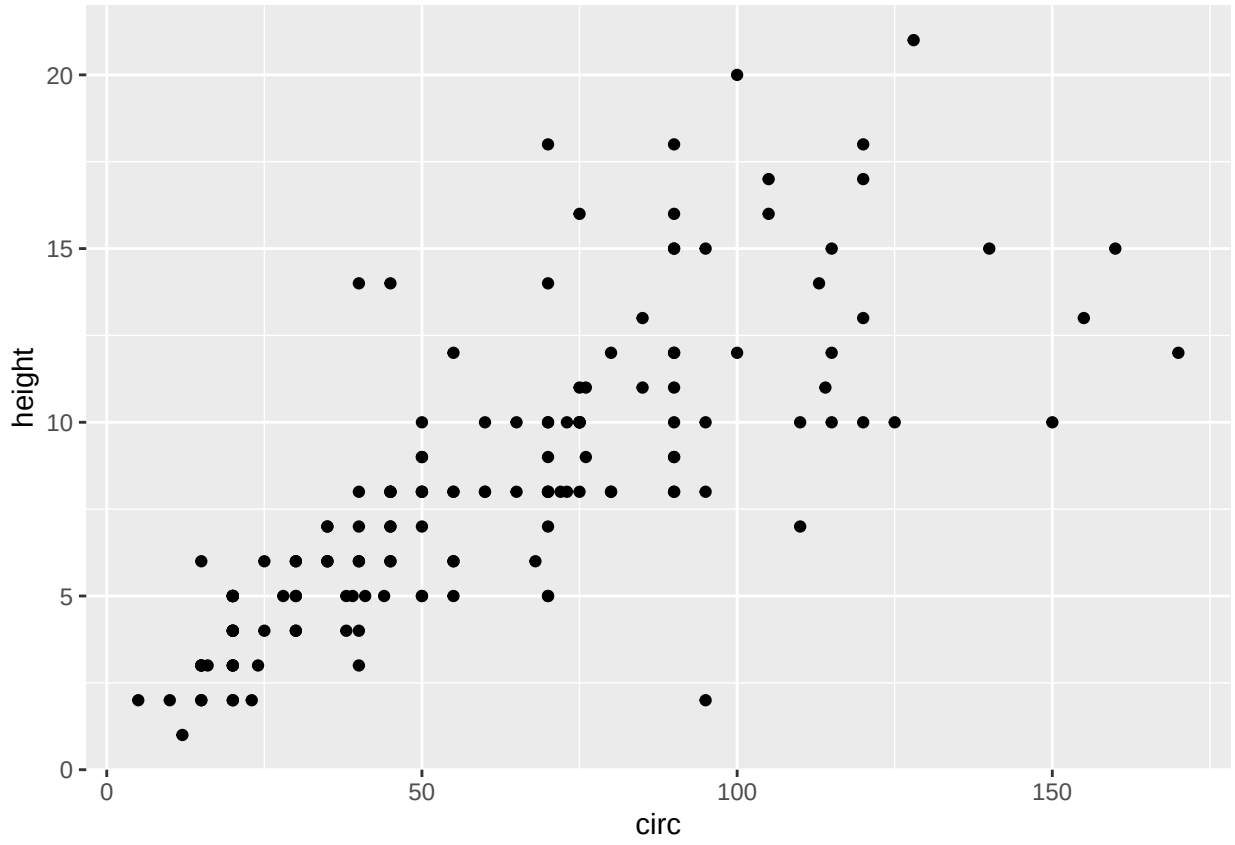
```
myData=read.table(file="arbres-tot.csv",sep=";",skip=3,header=TRUE)
myData=myData[myData$X10!=0,]
```

1. Simple regression

```
circ = myData$X70
height = myData$X10
simple_reg <- lm(height~circ,data=myData)
```

Plot the data to get an initial view of the possible correlation.

```
ggplot(myData,aes(x=circ,y=height))+ geom_point()+
xlab("circ")+
ylab("height")
```



```
anova(simple_reg)
```

```
## Analysis of Variance Table
##
## Response: height
##            Df Sum Sq Mean Sq F value    Pr(>F)
## circ        1 1508.1 1508.13  197.57 < 2.2e-16 ***
## Residuals 148 1129.8    7.63
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

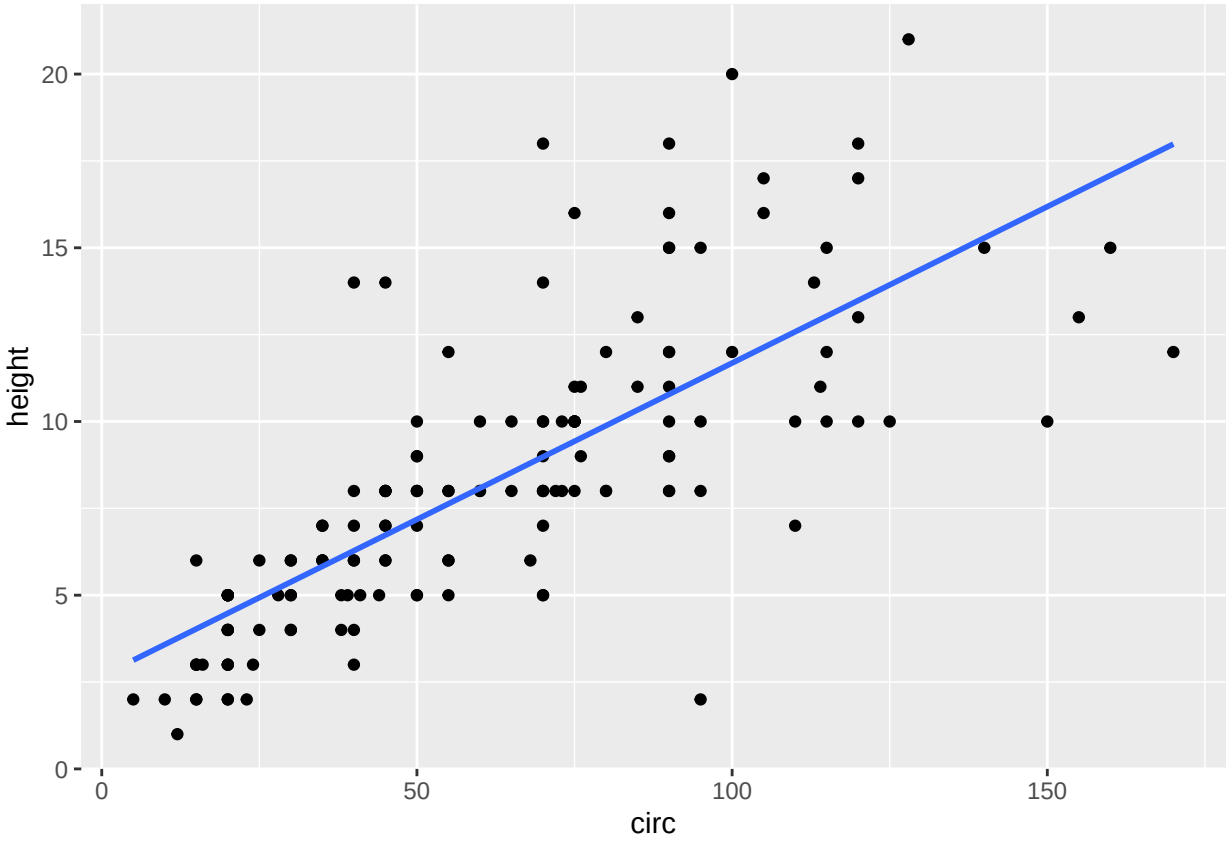
```
summary(simple_reg)
```

```
##
## Call:
## lm(formula = height ~ circ, data = myData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.2321 -1.6180 -0.2804  1.1280  9.0187
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.679057   0.455838   5.877 2.66e-08 ***
## circ        0.090032   0.006405  14.056  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.763 on 148 degrees of freedom
## Multiple R-squared:  0.5717, Adjusted R-squared:  0.5688
## F-statistic: 197.6 on 1 and 148 DF,  p-value: < 2.2e-16
```

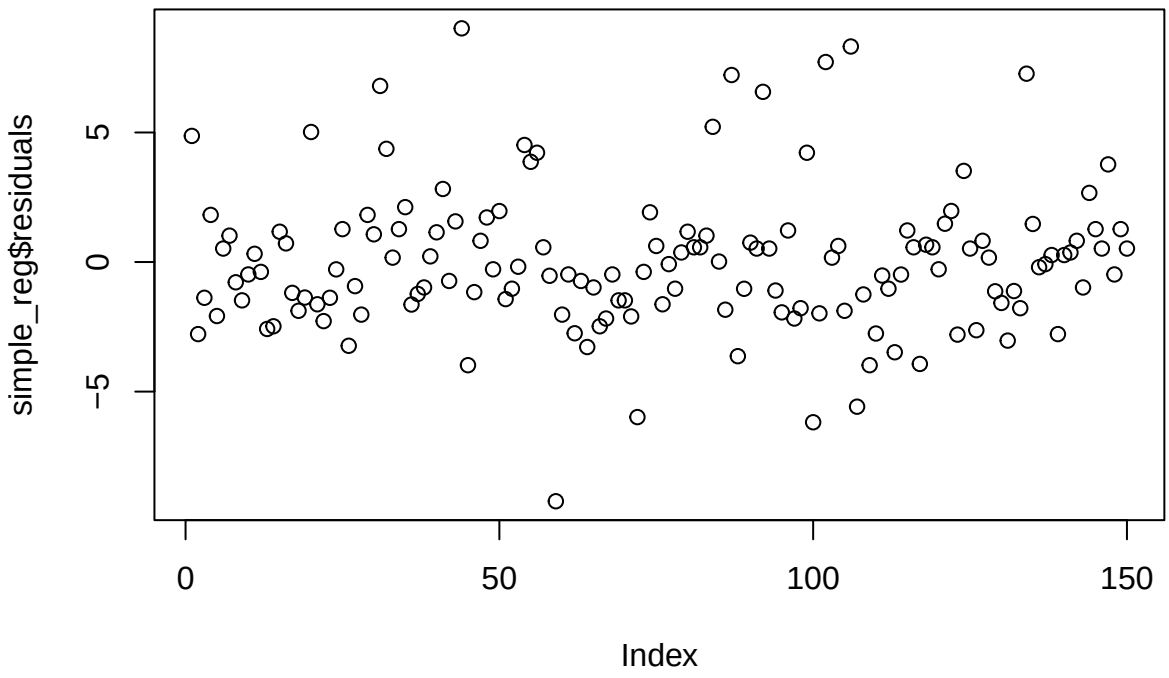
Plot the regression line

```
ggplot(myData,aes(x=circ,y=height))+ geom_point()+
stat_smooth(method="lm",se=FALSE)+ xlab("circ")+
ylab("height")
```

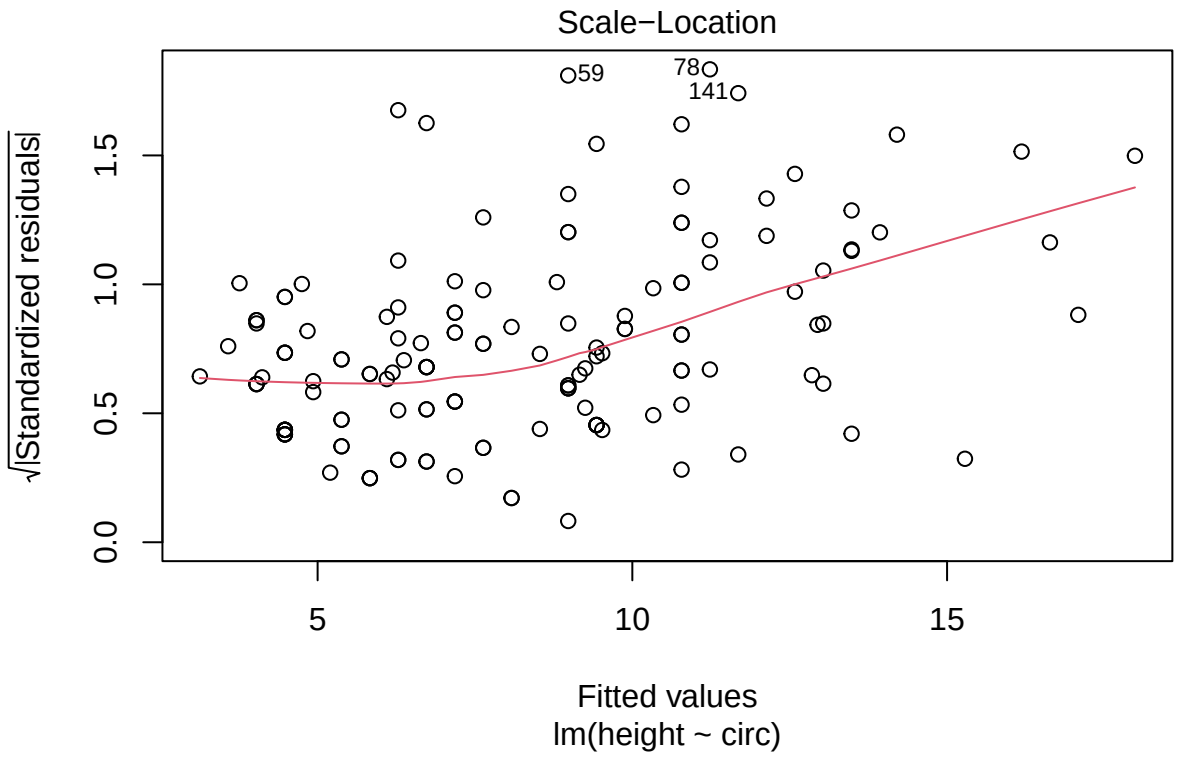
```
## `geom_smooth()` using formula 'y ~ x'
```



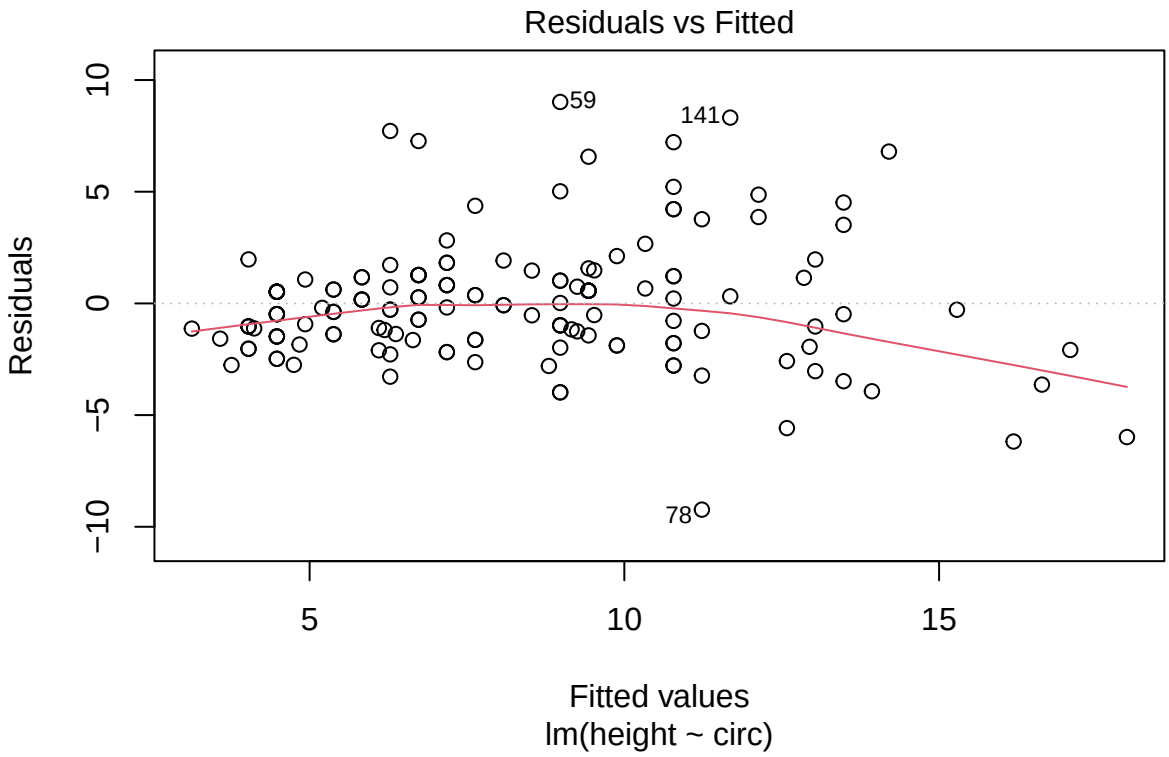

```
plot(simple_reg$residuals)
```



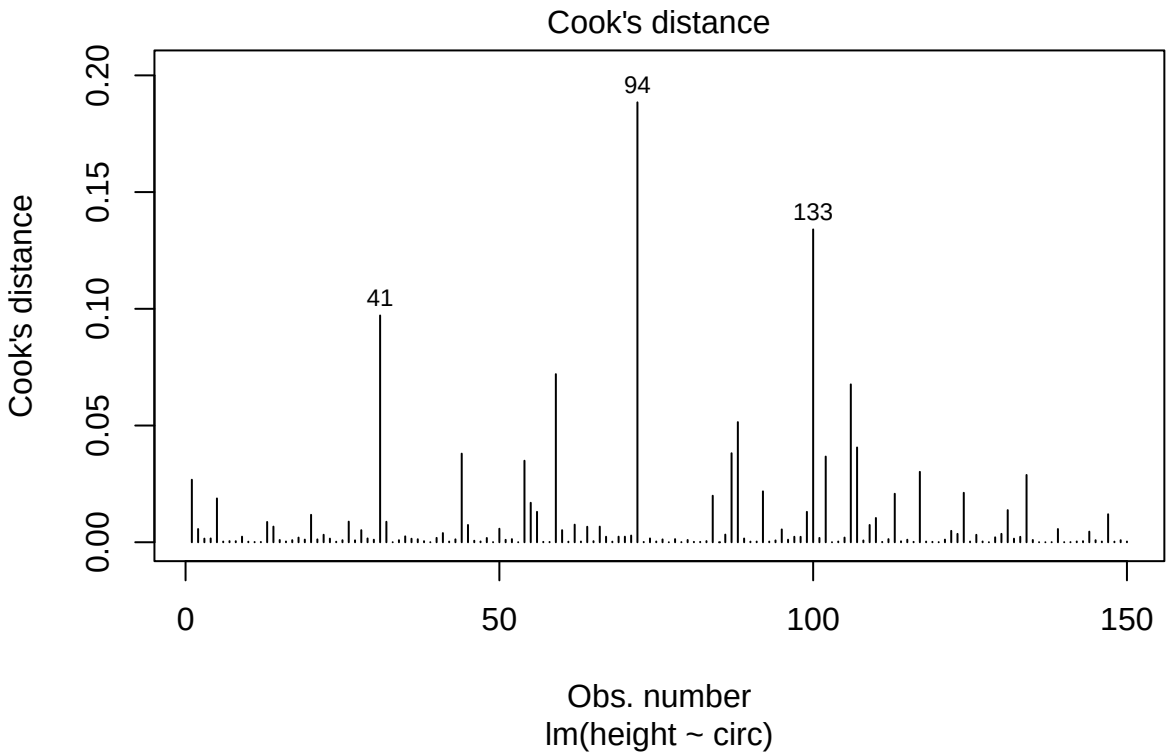

```
plot(simple_reg,3)
```




```
plot(simple_reg,1)
```




```
plot(simple_reg,4)
```




```
predict(simple_reg, data.frame(circ=10),interval="prediction")
```

```
##        fit       lwr     upr
## 1 3.579382 -1.937705 9.09647
```

2. Multivariate regression

Add a column to the sample. Name it circ_sqrt and fill it with the square root of the circumference of each tree

```
myData$circ_sqrt <- sqrt(myData$X70)
```

Perform the multivariate linear regression of height on the basis of: • of circumference ; • of circ_sqrt.

```
multi_reg <- lm(height~circ+circ_sqrt,data=myData)
summary(multi_reg)
```

```
##
## Call:
## lm(formula = height ~ circ + circ_sqrt, data = myData)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -9.4182 -1.5795 -0.0383  0.9617  8.4205
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.96947    2.05237  -1.934  0.05502 .
## circ        -0.02947    0.03656  -0.806  0.42149
## circ_sqrt    1.86596    0.56255   3.317  0.00115 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.674 on 147 degrees of freedom
## Multiple R-squared:  0.6015, Adjusted R-squared:  0.5961
## F-statistic:   111 on 2 and 147 DF,  p-value: < 2.2e-16
```

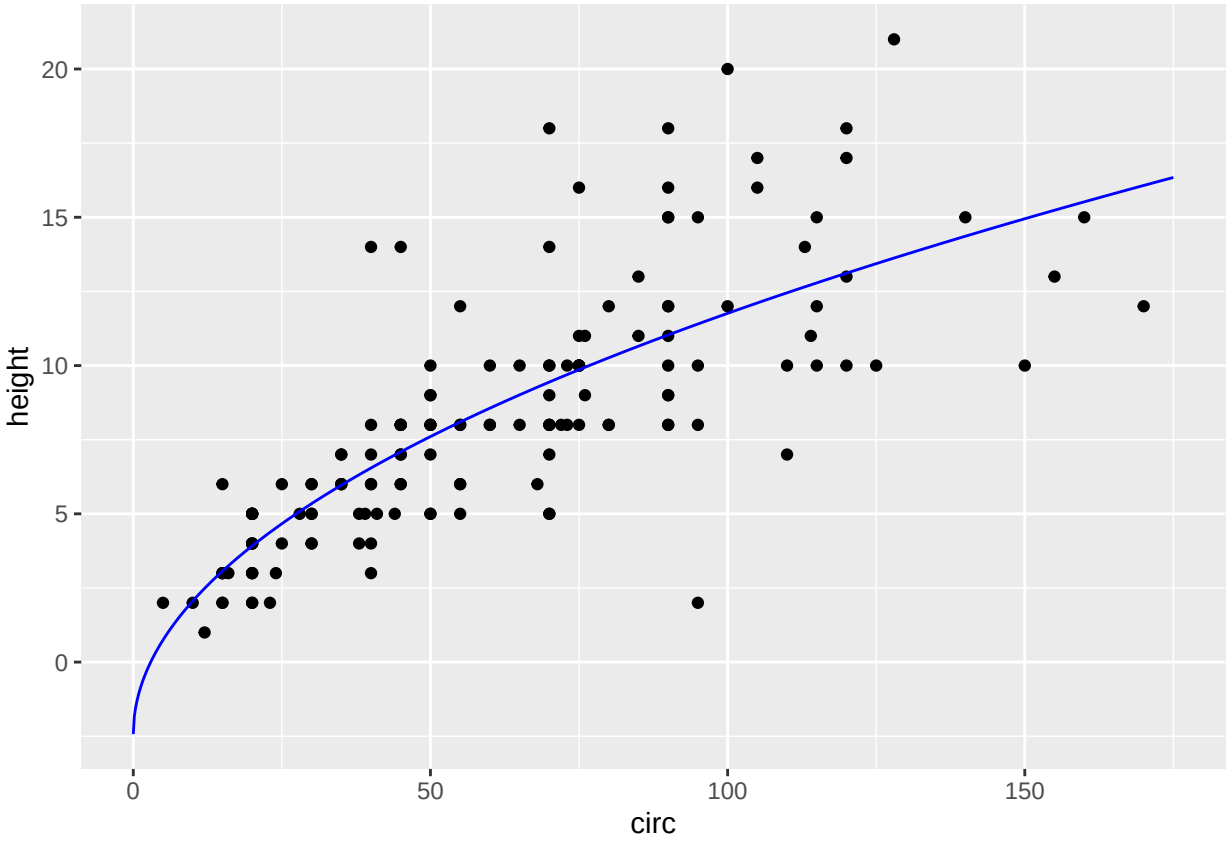
The variable **circ** is not significant at the 5% test level, so we will remove it for the regression.

```
multi_reg_2 <- lm(height~circ_sqrt,data=myData)
summary(multi_reg_2)
```

```
##
## Call:
## lm(formula = height ~ circ_sqrt, data = myData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.4017 -1.5133 -0.0729  1.0343  8.5568
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.42957    0.74930  -3.242  0.00146 **
## circ_sqrt    1.41906    0.09528  14.893  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.671 on 148 degrees of freedom
## Multiple R-squared:  0.5998, Adjusted R-squared:  0.5971
## F-statistic: 221.8 on 1 and 148 DF,  p-value: < 2.2e-16
```

Plot the regression

```
circ_pred <- seq(0,175,len=1000)
height_pred <- multi_reg_2$coefficients[1]+multi_reg_2$coefficients[2]*sqrt(circ_pred)
fct_reg <- data.frame(circ_pred=circ_pred,height_pred=height_pred)
ggplot() + geom_point(data=myData,aes(x=circ,y=height)) + geom_line(data=fct_reg,aes(x=circ_pred,y=heigh
```


height
circ
0
5
10
15
20
0
50
100
150