Assignment 7

Mandi Bluth

2024-11-08

Exercise 2

Using the `datasets::trees` data, complete the following. This question refreshes create a linear model, graphing the linear model, and introduces using some LaTeX expressions on the graph.

a) Create a regression model for $y =$ `Volume` as a function of $x =$ `Height`.

```
data(trees)
trees.lm <- lm(data=trees, Volume ~ Height)
```

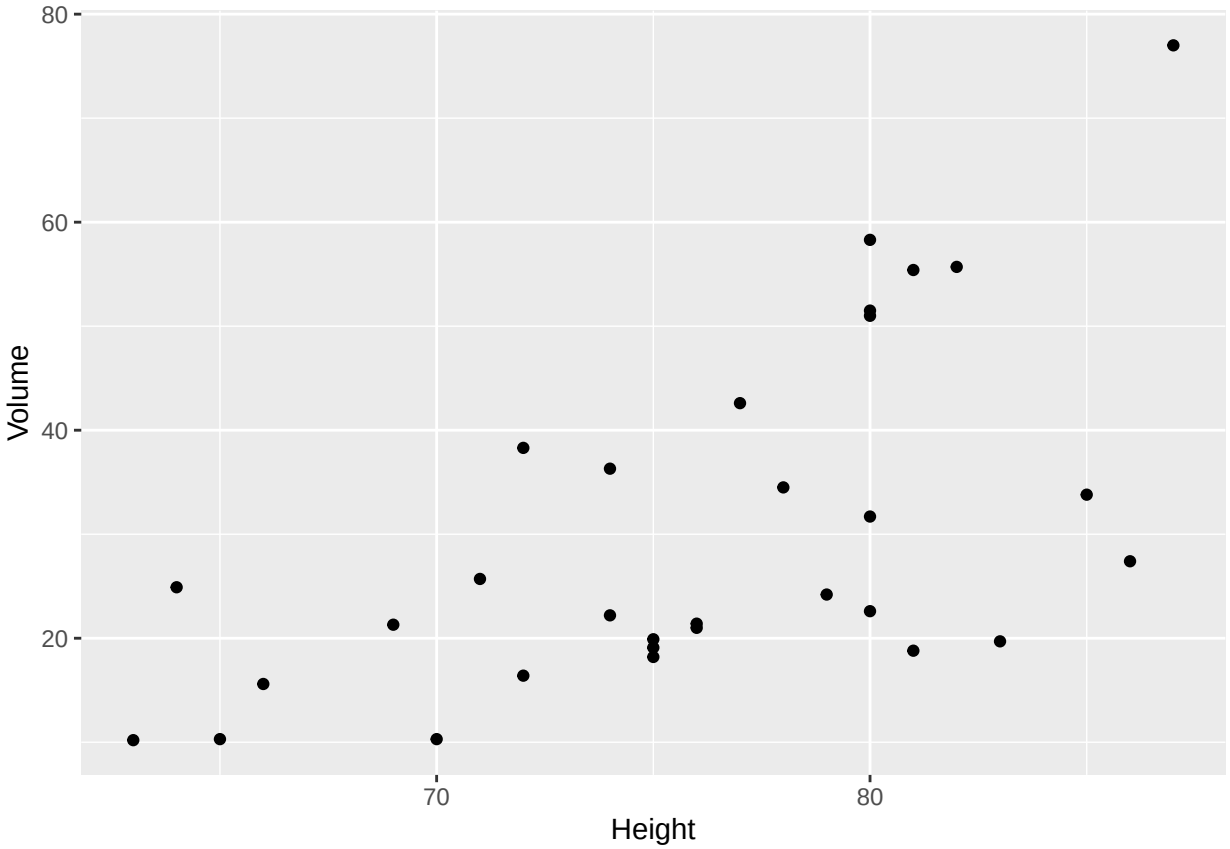
b) Display the `summary` of the model to view the y-intercept and slope of the regression line.

```
summary(trees.lm)
```

```
##
## Call:
## lm(formula = Volume ~ Height, data = trees)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.274  -9.894  -2.894  12.068  29.852
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -87.1236    29.2731  -2.976 0.005835 **
## Height        1.5433     0.3839   4.021 0.000378 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.4 on 29 degrees of freedom
## Multiple R-squared:  0.3579, Adjusted R-squared:  0.3358
## F-statistic: 16.16 on 1 and 29 DF,  p-value: 0.0003784
```

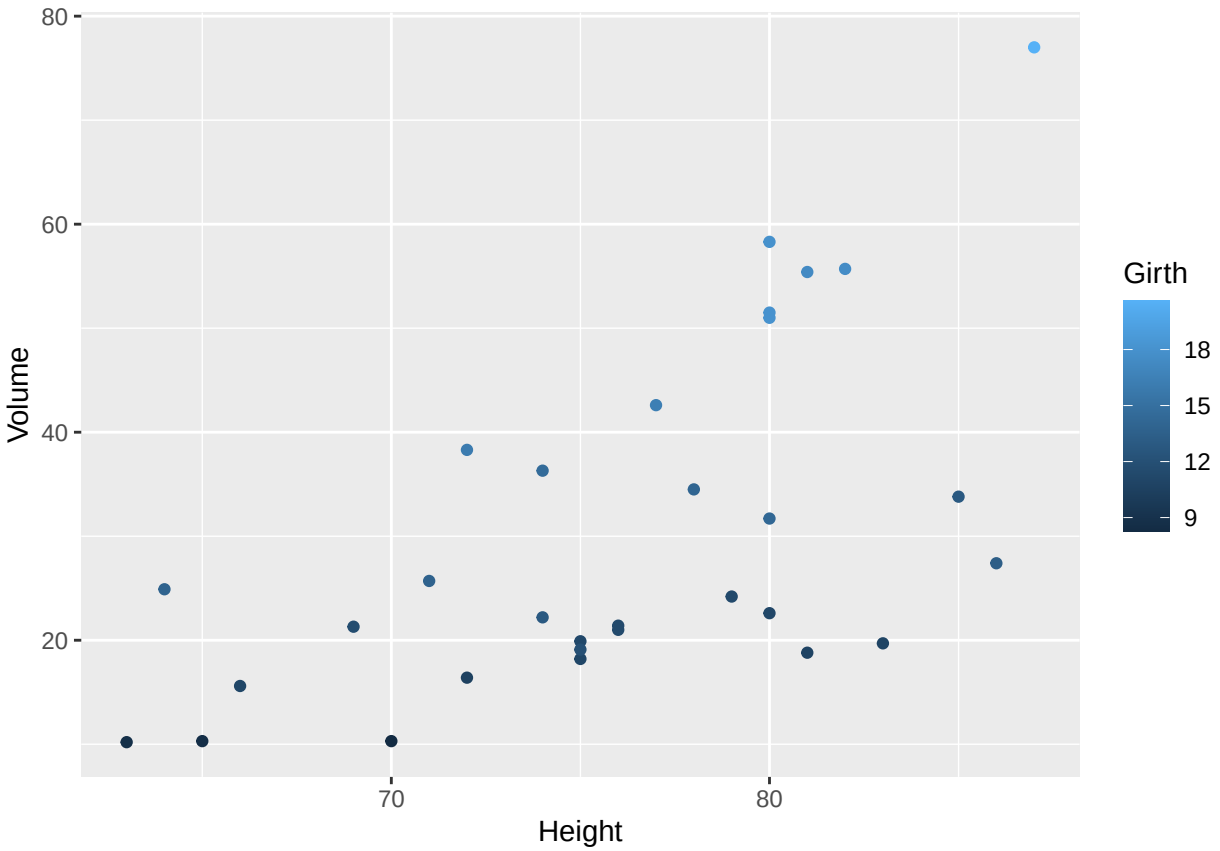
c) Using `ggplot2`, create a scatter plot of Volume vs Height.

```
ggplot(data=trees ,
mapping=aes(x=Height , y=Volume))+
geom_point()
```



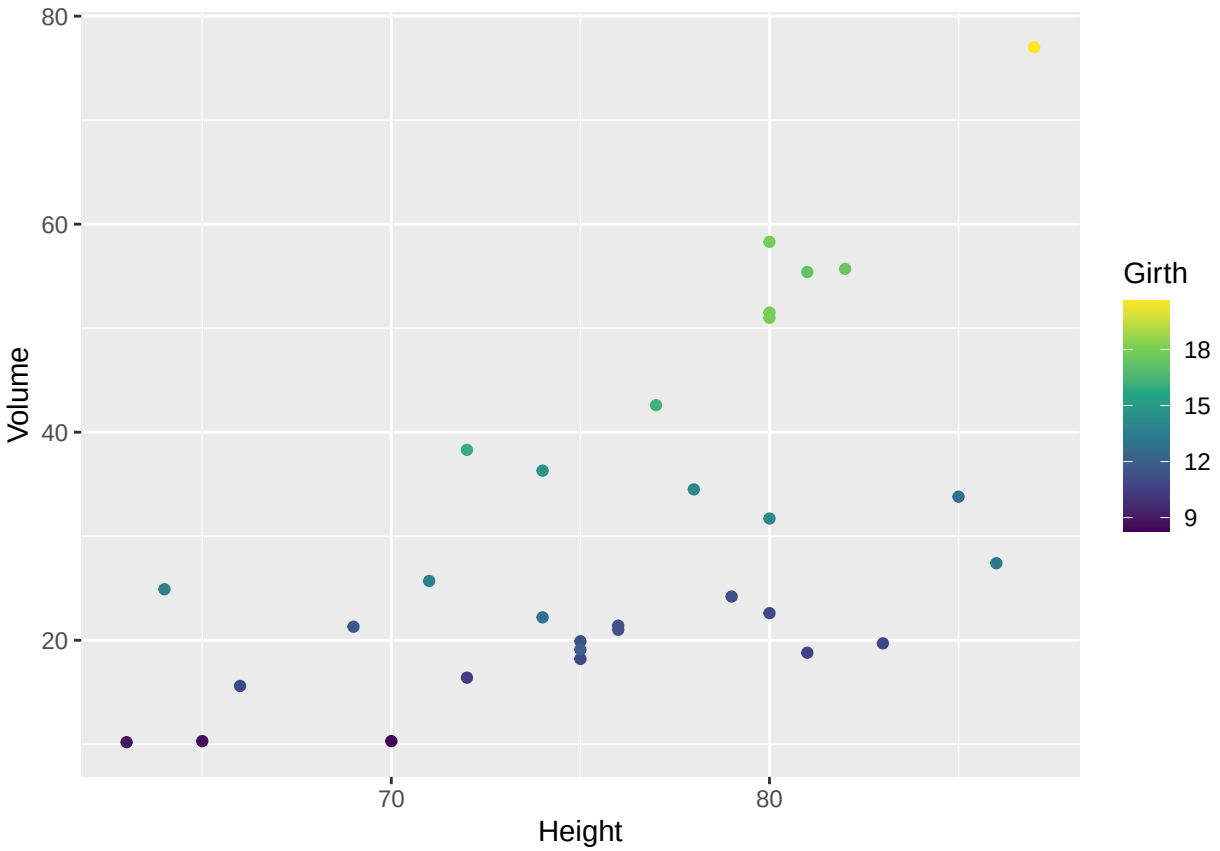
d) Color the scatter using the `Girth` variable.

```
ggplot(data=trees ,
mapping=aes(x=Height , y=Volume, color=Girth))+
geom_point()
```



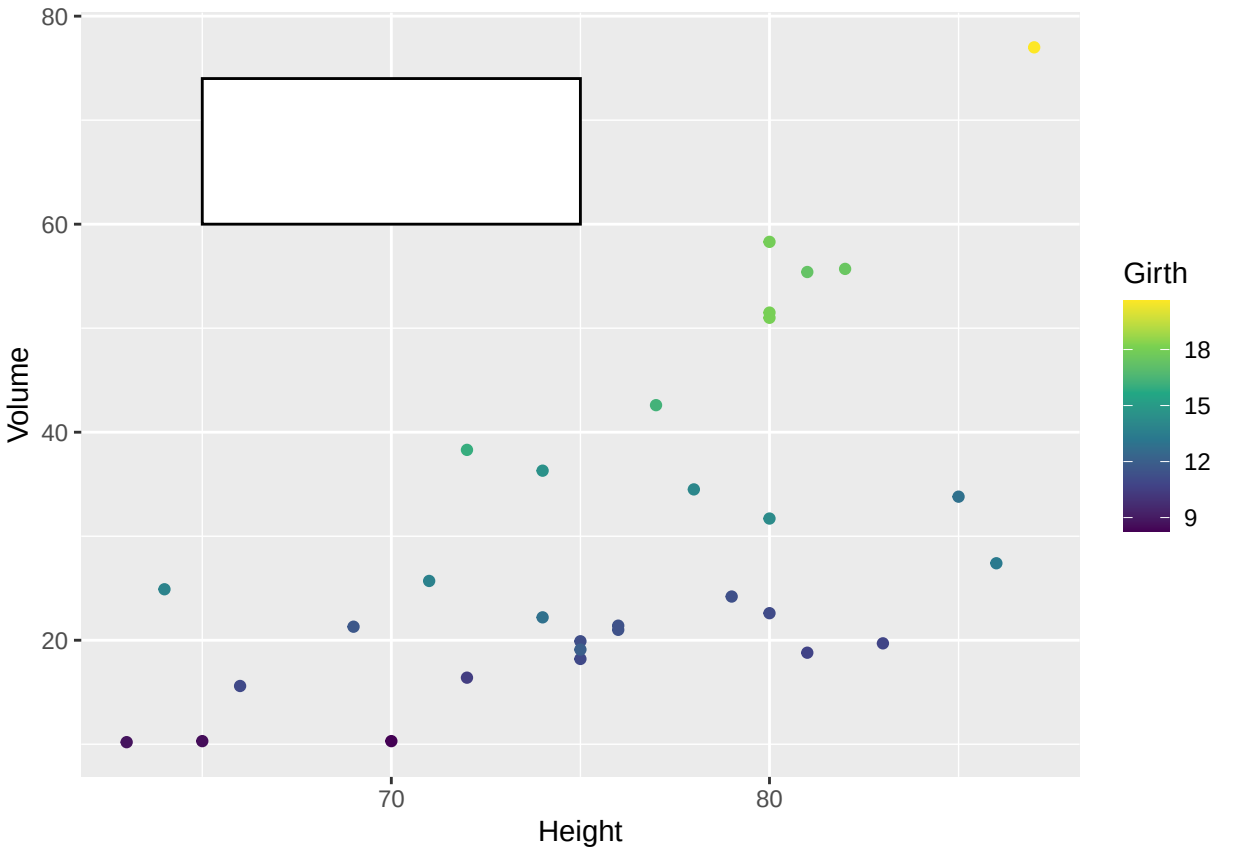
e) Modify the color scheme using a RColorBrewer palette.

```
ggplot(data=trees ,
mapping=aes(x=Height , y=Volume, color=Girth))+
geom_point()+
  scale_color_viridis_c(option='viridis')
```



f) Create a nice white filled rectangle to add text information. The following might be useful.

```
ggplot(data=trees ,
mapping=aes(x=Height , y=Volume, color=Girth))+
geom_point() + scale_color_viridis_c(option='viridis') +
annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74, fill='white', color='black')
```

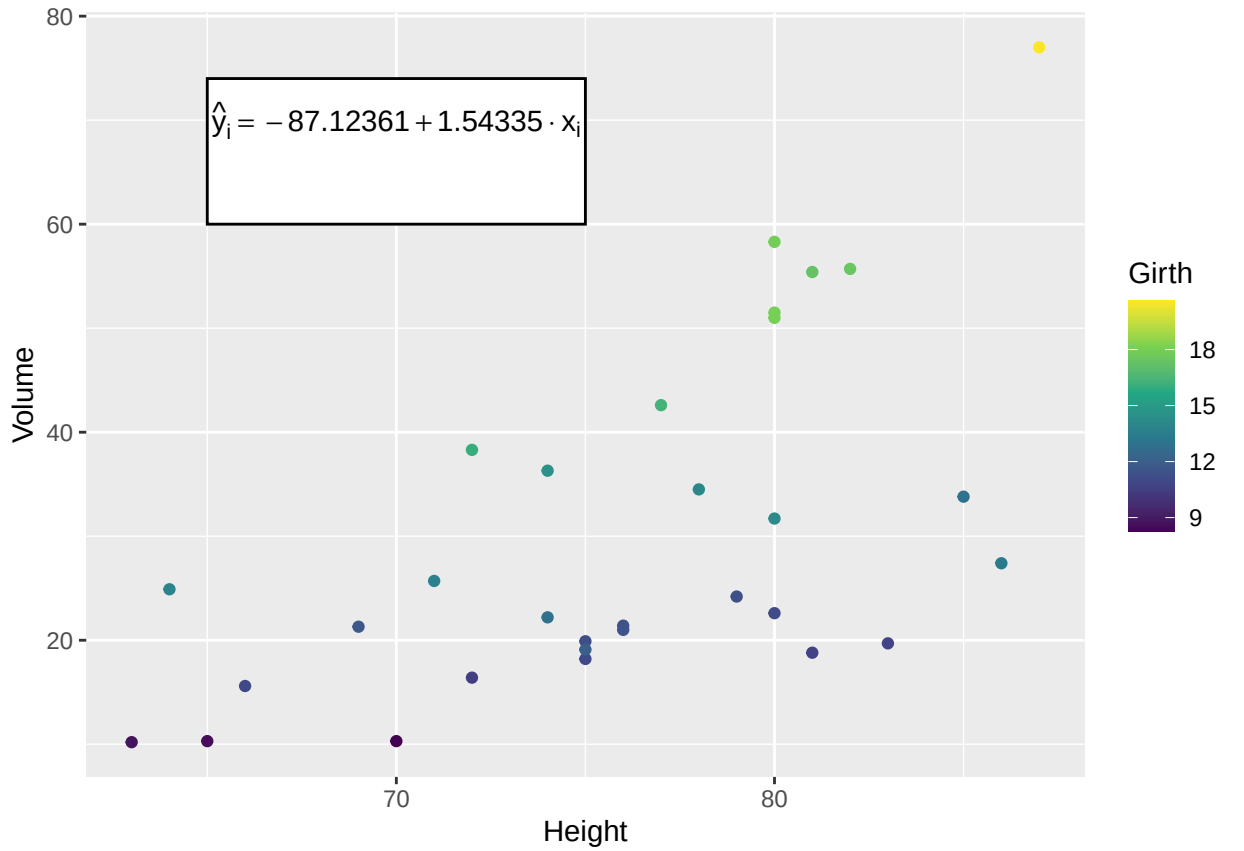


g) Use the `broom` package to extract the coefficients of the best-fit line. Add this information as an annotation to the graph, which should follow a form that looks like $\hat{y}_i = (INTERCEPT) + (SLOPE) * x_i$. Place the annotation within the white text box.

```
tidy(trees.lm)
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)   -87.1     29.3       -2.98 0.00583
## 2 Height          1.54     0.384      4.02 0.000378
```

```
ggplot(data=trees ,
mapping=aes(x=Height , y=Volume, color=Girth))+
geom_point() + scale_color_viridis_c(option='viridis') +
annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
         fill='white', color='black')+
  annotate('text', x=70, y=70, size = 4,
           label = TeX('$\\hat{y}_i = -87.12361 + 1.54335 \\cdot x_i$'))
```

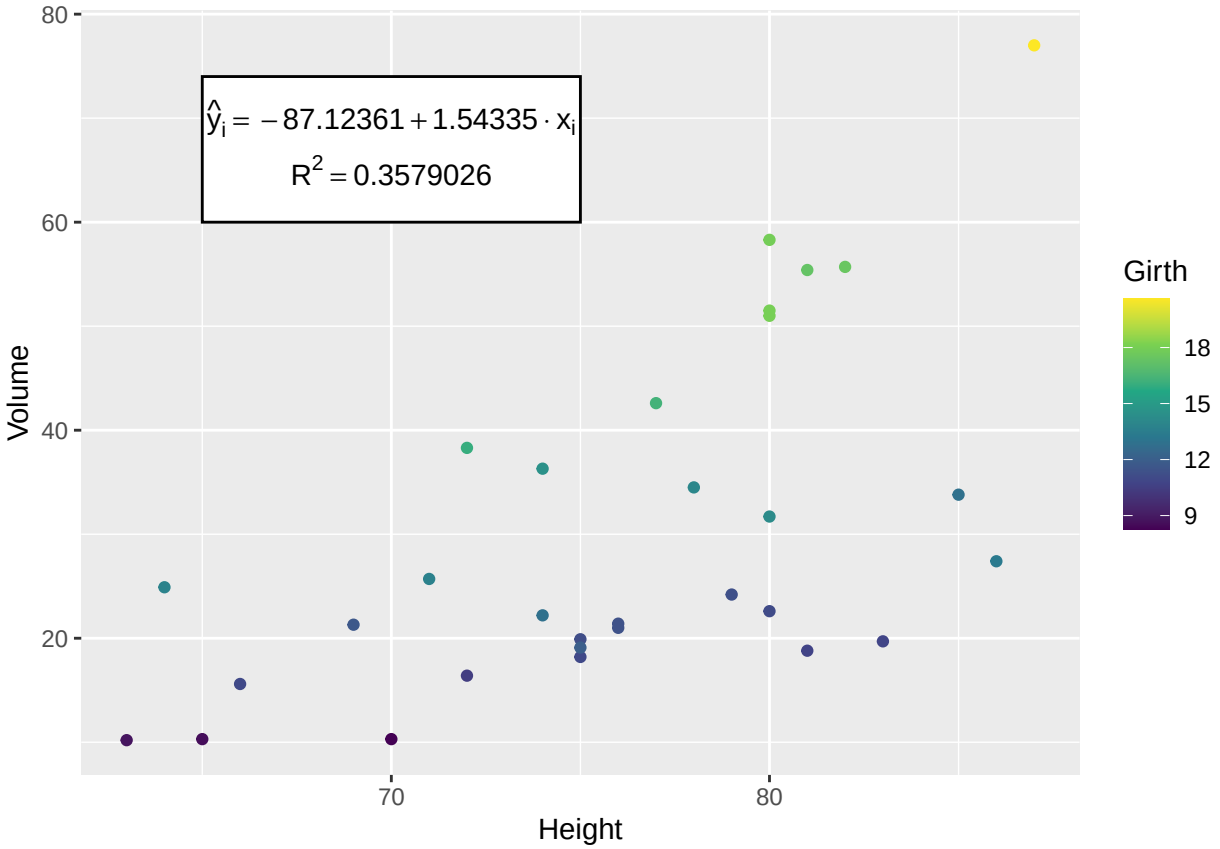


h) Use the `broom` package to extract the coefficient of determination r^2 from the model. Add the annotation to your graph, which should look something like $R^2 = (VALUE)$

```
summary(trees.lm)$r.squared
```

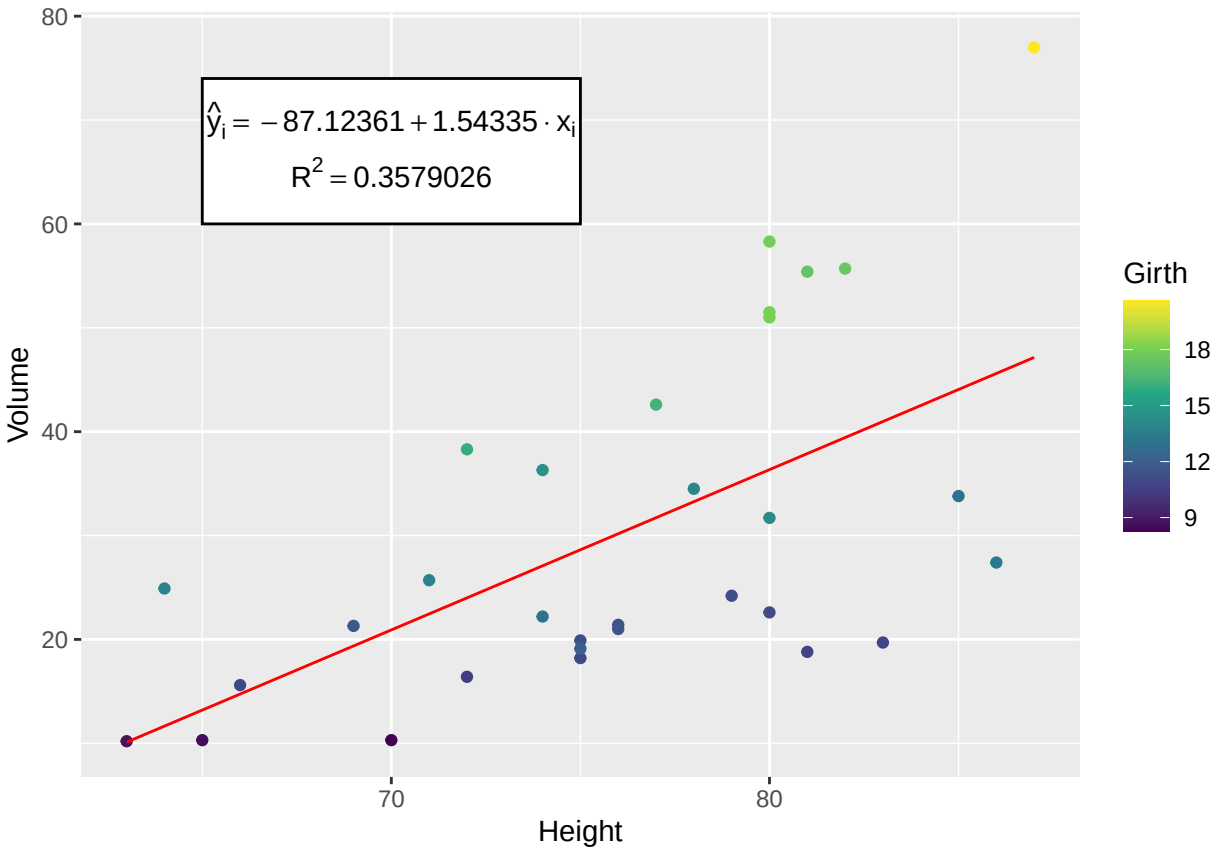
```
## [1] 0.3579026
```

```
ggplot(data=trees ,
mapping=aes(x=Height , y=Volume, color=Girth))+
geom_point() + scale_color_viridis_c(option='viridis') +
annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
         fill='white', color='black')+
  annotate('text', x=70, y=70, size = 4,
           label = TeX('$\\hat{y}_i = -87.12361 + 1.54335 \\cdot x_i$')) +
  annotate('text', x=70, y=65, size = 4,
           label = TeX('$R^2 = 0.3579026$'))
```



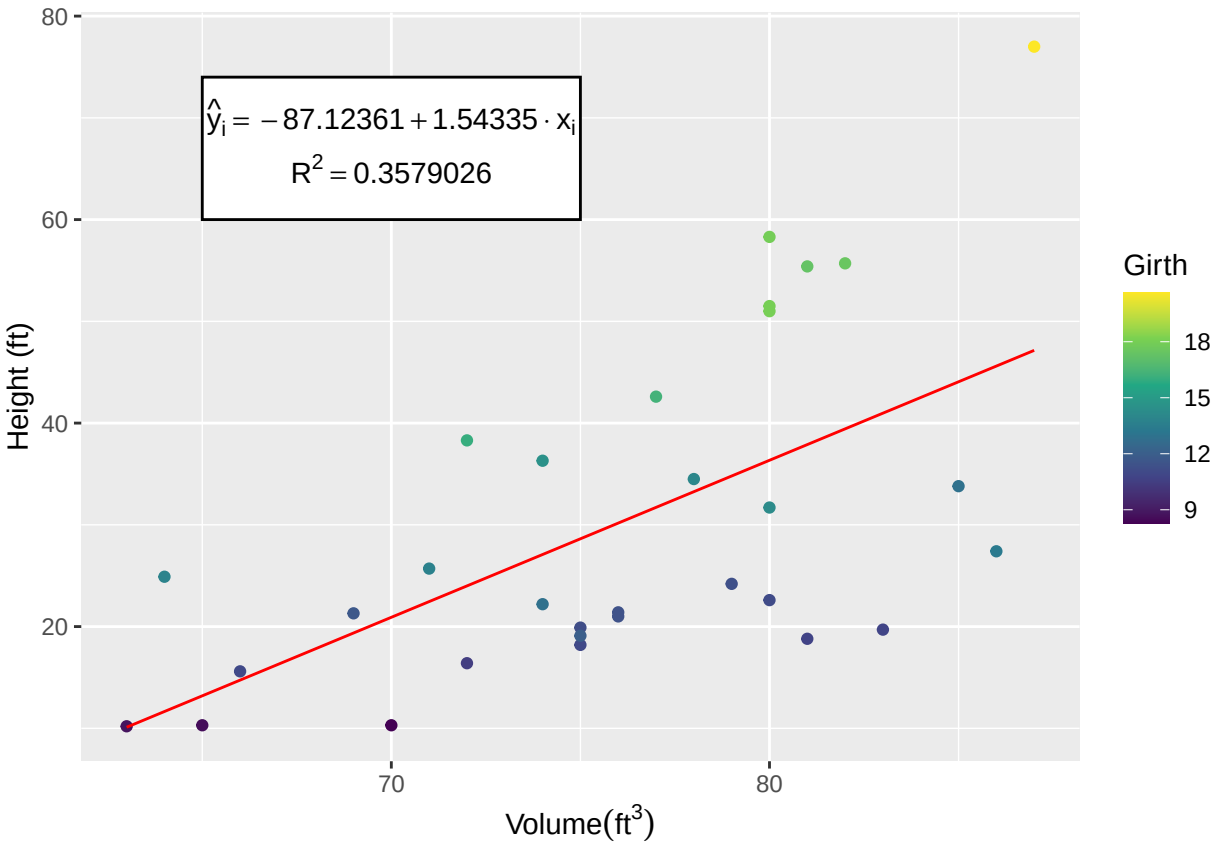
i) Add the regression line in red. There are several ways to do this.

```
trees <- trees %>%
  mutate(fit=fitted(trees.lm))
ggplot(data=trees ,
mapping=aes(x=Height , y=Volume, color=Girth))+
geom_point() + scale_color_viridis_c(option='viridis') +
annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
         fill='white', color='black')+
  annotate('text', x=70, y=70, size = 4,
           label = TeX('$\\hat{y}_i = -87.12361 + 1.54335 \\cdot x_i$')) +
  annotate('text', x=70, y=65, size = 4,
           label = TeX('$R^2 = 0.3579026$'))+
  geom_line(aes(y=fit), color='red')
```



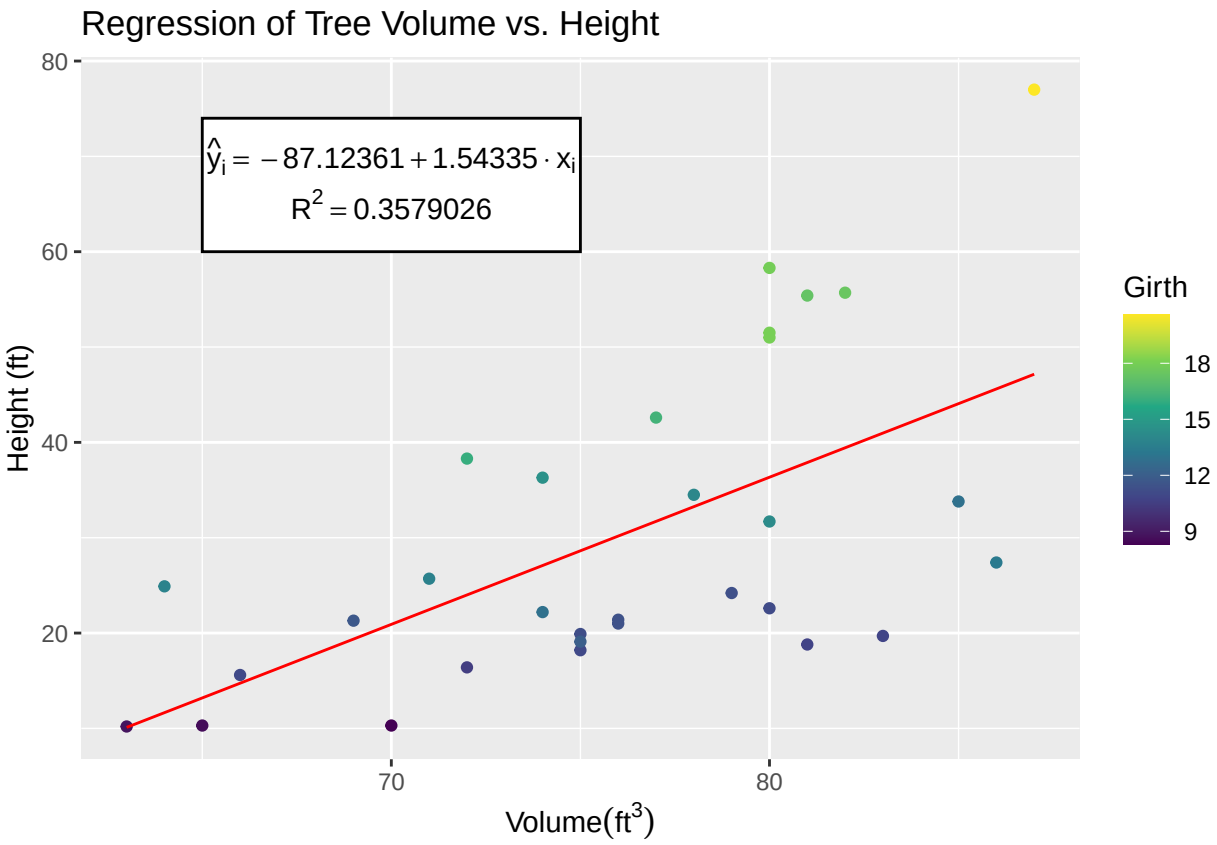
j) Properly label the axes of the graph.

```
ggplot(data=trees ,
mapping=aes(x=Height , y=Volume, color=Girth))+
geom_point() + scale_color_viridis_c(option='viridis') +
annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
         fill='white', color='black')+
  annotate('text', x=70, y=70, size = 4,
           label = TeX('$\\hat{y}_i = -87.12361 + 1.54335 \\cdot x_i$')) +
  annotate('text', x=70, y=65, size = 4, label = TeX('$R^2 = 0.3579026$'))+
  geom_line(aes(y=fit), color='red')+
  labs(x = TeX("$Volume (ft^3)$"), y = "Height (ft)")
```



k) Add a descriptive title to the graph.

```
ggplot(data=trees ,
mapping=aes(x=Height , y=Volume, color=Girth))+
geom_point() + scale_color_viridis_c(option='viridis') +
annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
         fill='white', color='black')+
  annotate('text', x=70, y=70, size = 4,
           label = TeX('$\\hat{y}_i = -87.12361 + 1.54335 \\cdot x_i$')) +
  annotate('text', x=70, y=65, size = 4,
           label = TeX('$R^2 = 0.3579026$'))+
  geom_line(aes(y=fit), color='red')+
  labs(x = TeX("$Volume (ft^3)$"), y = "Height (ft)",
       title = "Regression of Tree Volume vs. Height")
```


Regression of Tree Volume vs. Height
$\hat{y}_i = -87.12361 + 1.54335 \cdot x_i$
$R^2 = 0.3579026$
Height (ft)
Volume(ft^3)
80
60
40
20
70
80
Girth
18
15
12
9