Assignment 7

Landon Wilson

2024-11-08

```
library(tidyverse)
library(viridis)
library(latex2exp)
library(plotly)
library(RColorBrewer)
```

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`.

```
trees.model <- lm(trees$Volume ~ trees$Height)
```

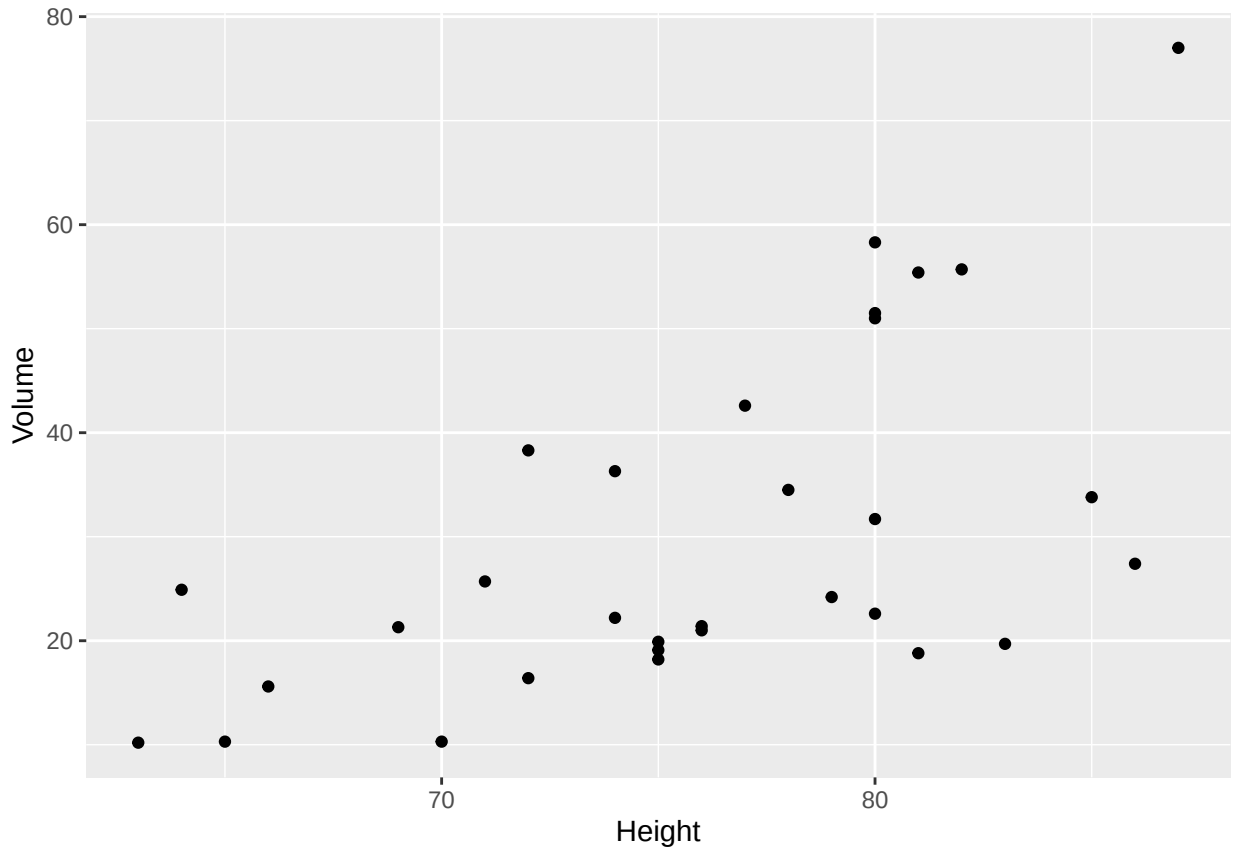
b) Display the `summary` of the model to view the y-intercept and slope of the regression line.

```
summary(trees.model)
```

```
##
## Call:
## lm(formula = trees$Volume ~ trees$Height)
##
## 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 **
## trees$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
```

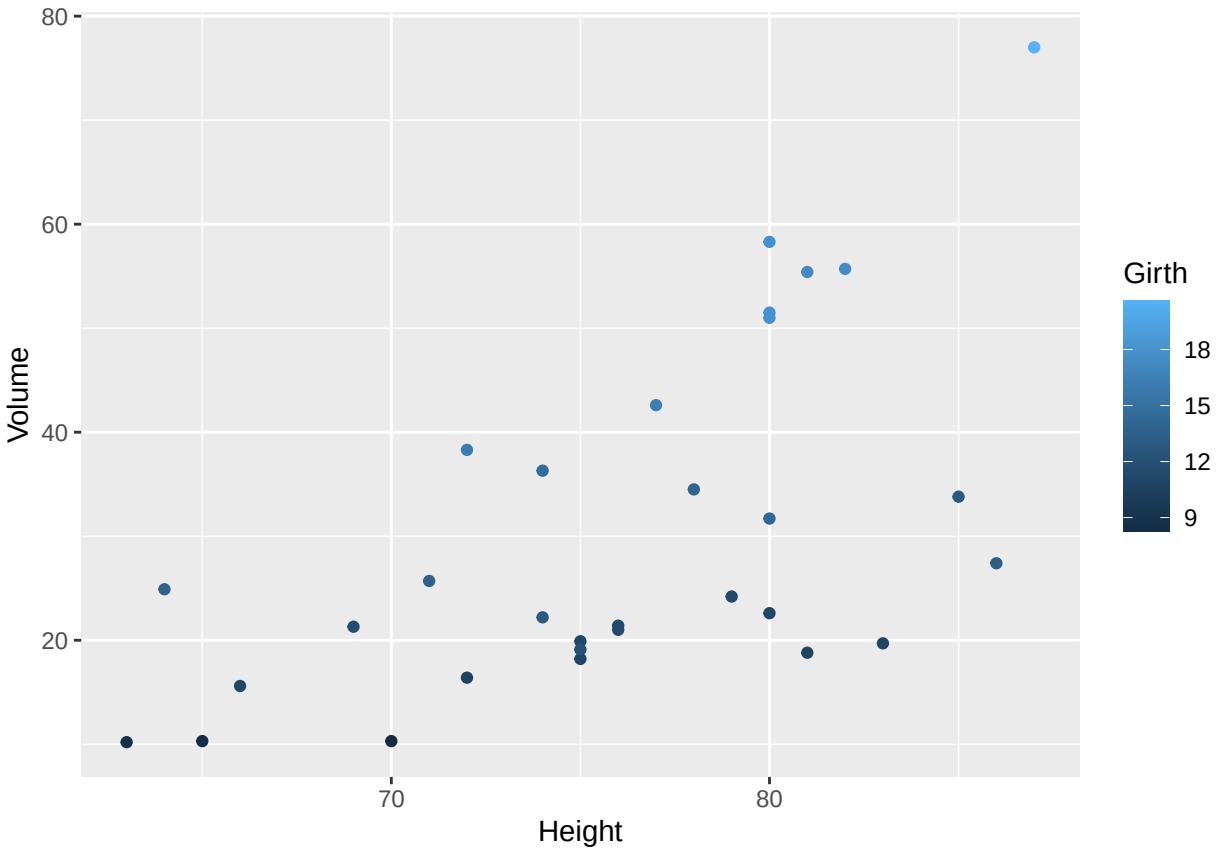
c) Using `ggplot2`, create a scatter plot of Volume vs Height.

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



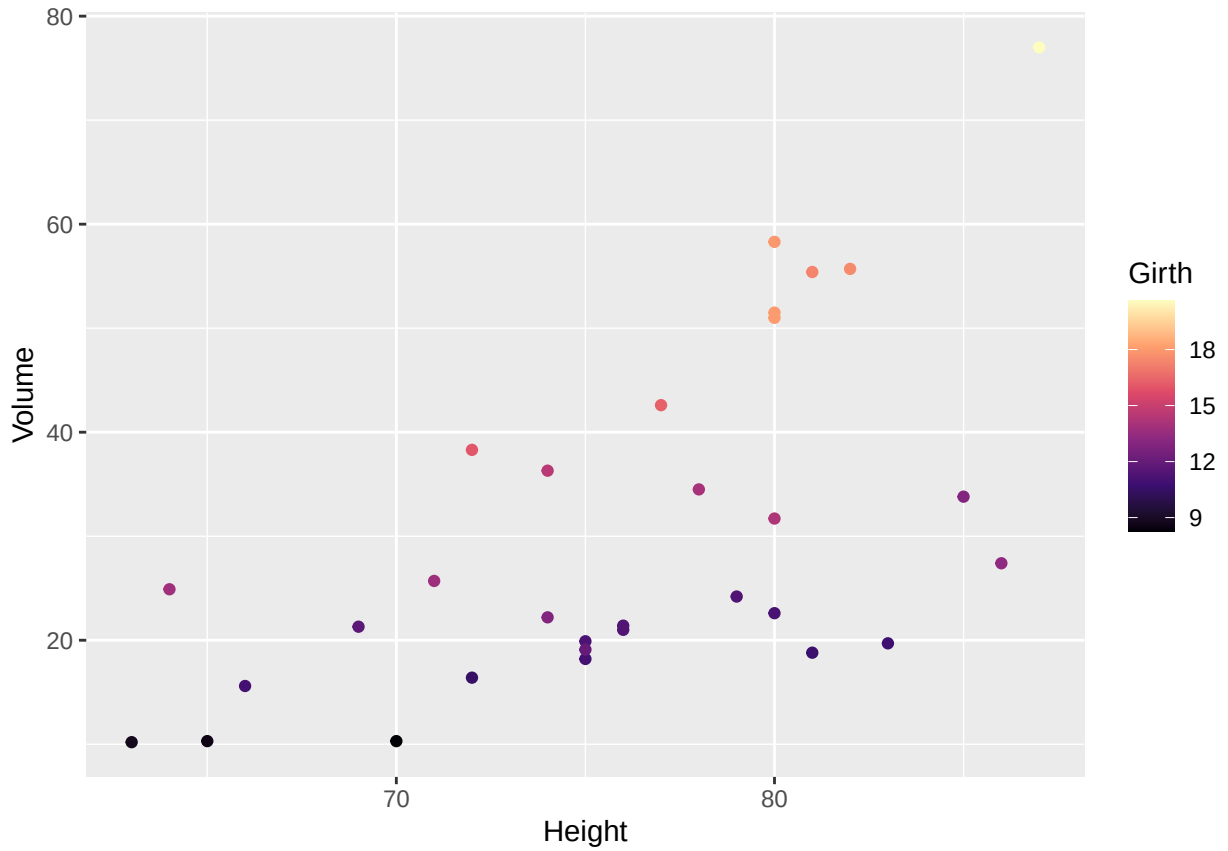
d) Color the scatter using the `Girth` variable.

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



e) Modify the color scheme using a RColorBrewer palette.

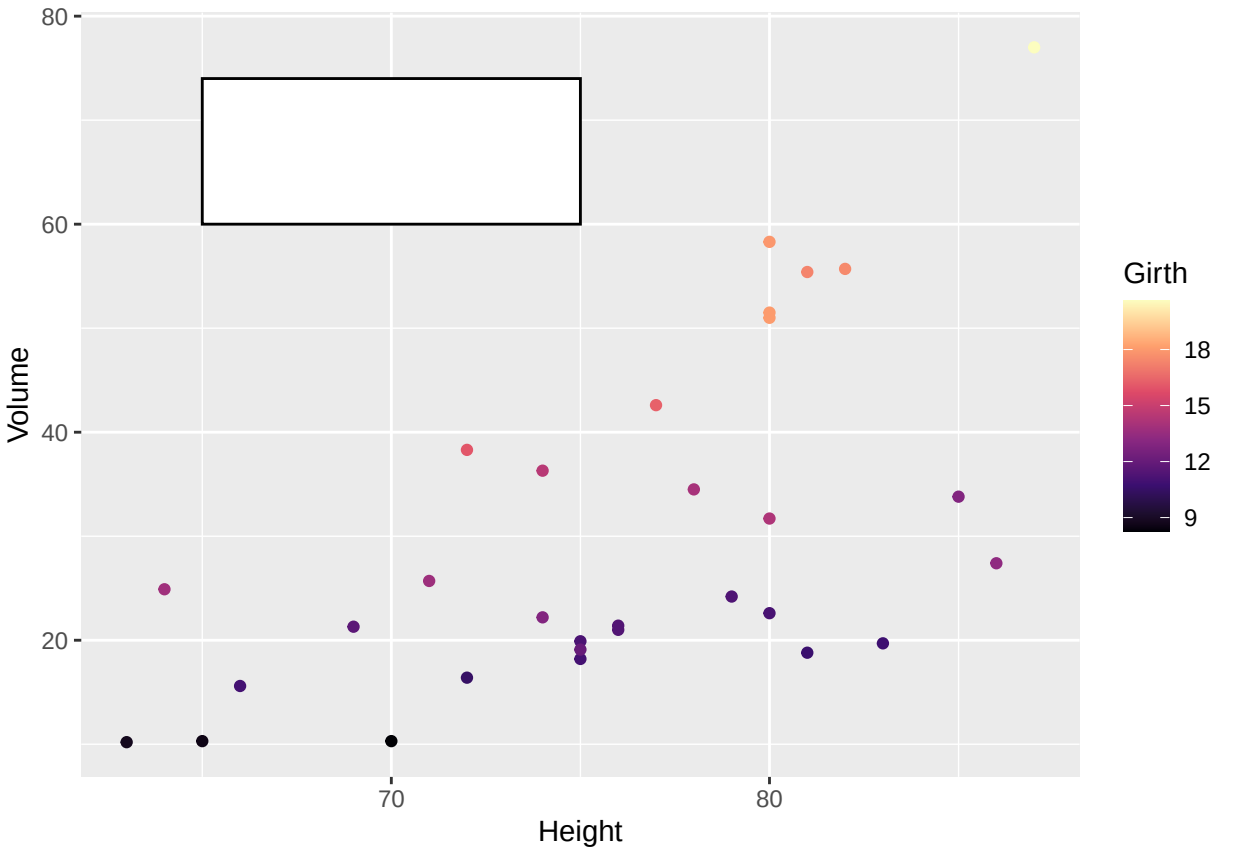
```
ggplot(data = trees, aes(x = Height, y = Volume)) +
  geom_point(aes(color = Girth)) +
  scale_color_viridis_c(option='magma')
```



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

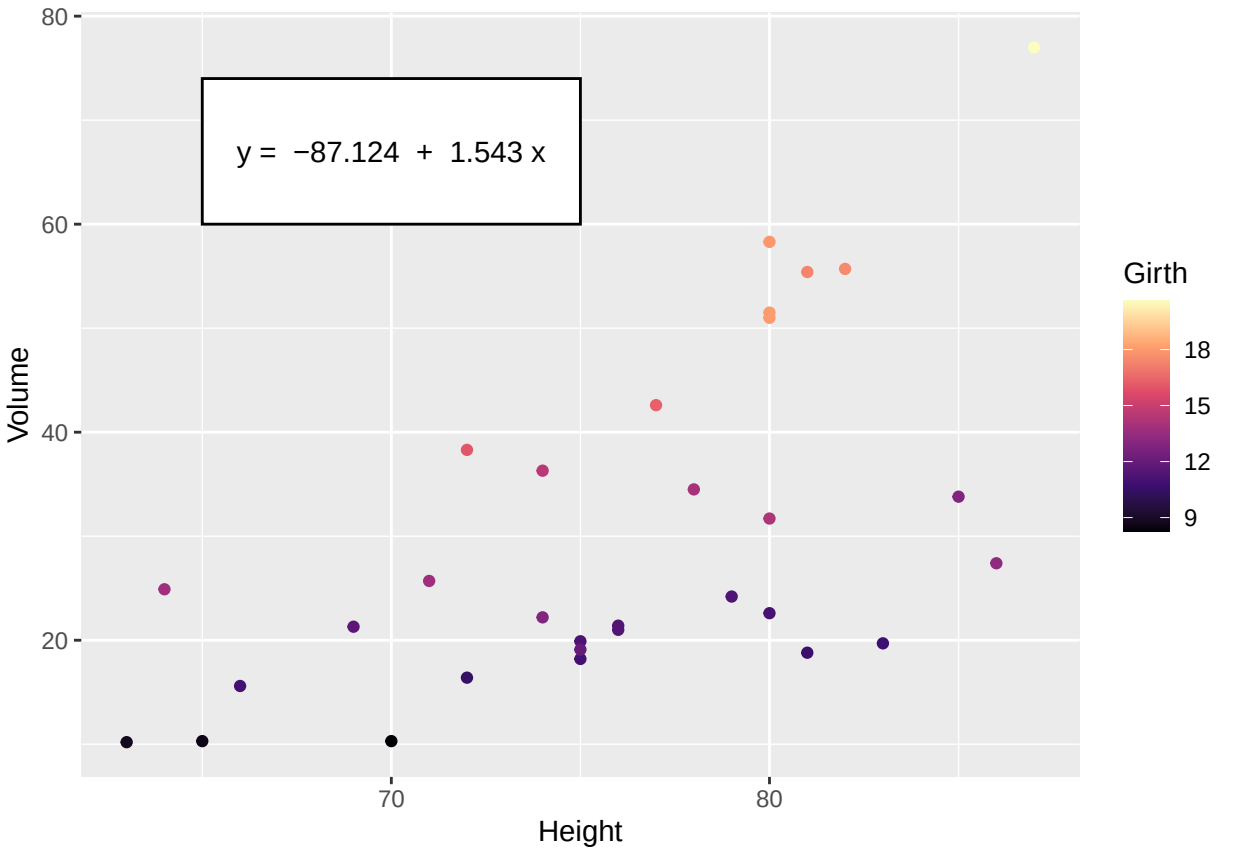
```
 annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
         fill='white', color='black')+
```

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



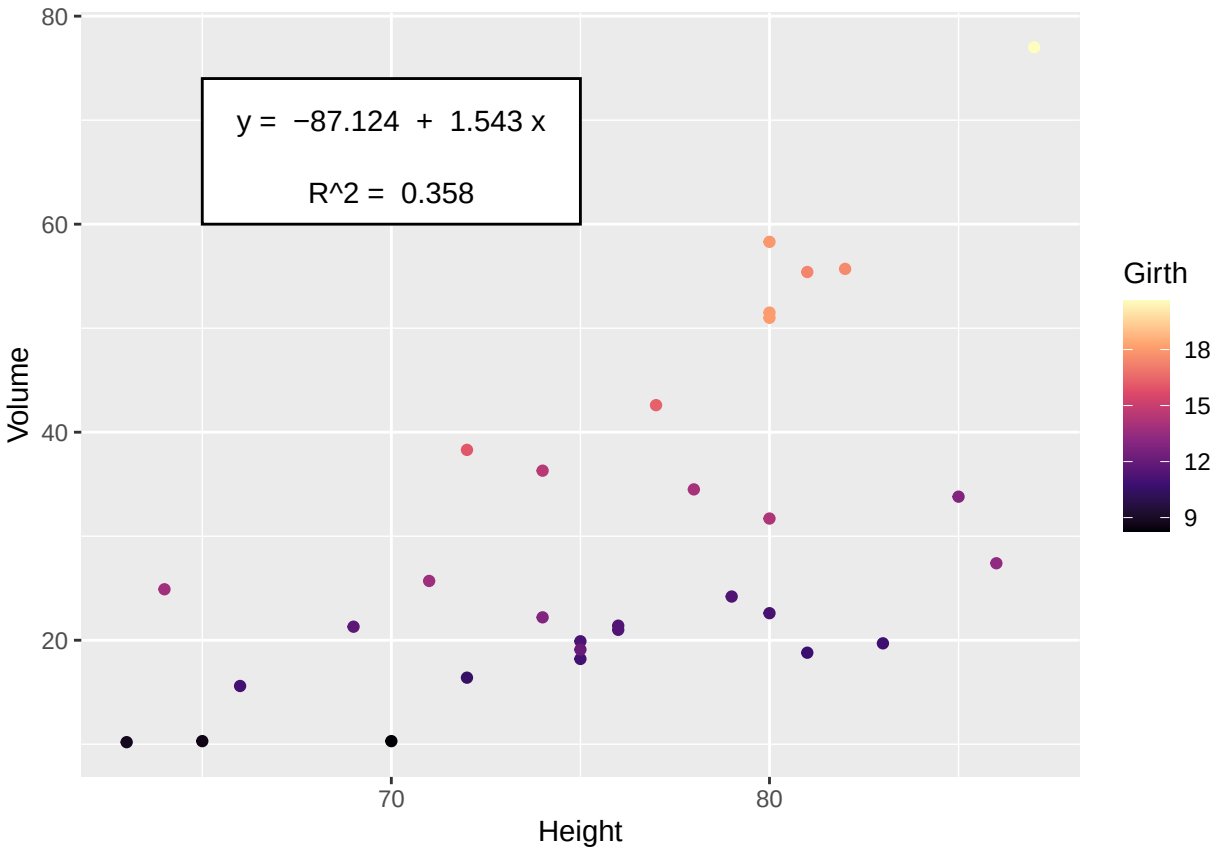
g) Use the `broom` package to extract the coefficients of the best-fit line. Add this information as an annotation to the graph, which should follow a form that looks like $\hat{y}_i = (INTERCEPT) + (SLOPE) * x_i$. Place the annotation within the white text box.

```
coef <- broom::tidy(trees.model)
int <- round(coef[1,2],3)
slope <- round(coef[2,2],3)

ggplot(data = trees, aes(x = Height, y = Volume)) +
  geom_point(aes(color = Girth)) +
  scale_color_viridis_c(option='magma')+
  annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
         fill='white', color='black')+
  annotate("text", x = 70, y = 67,
           label = paste("y = ", int, " + ", slope, "x"), parse = FALSE)
```



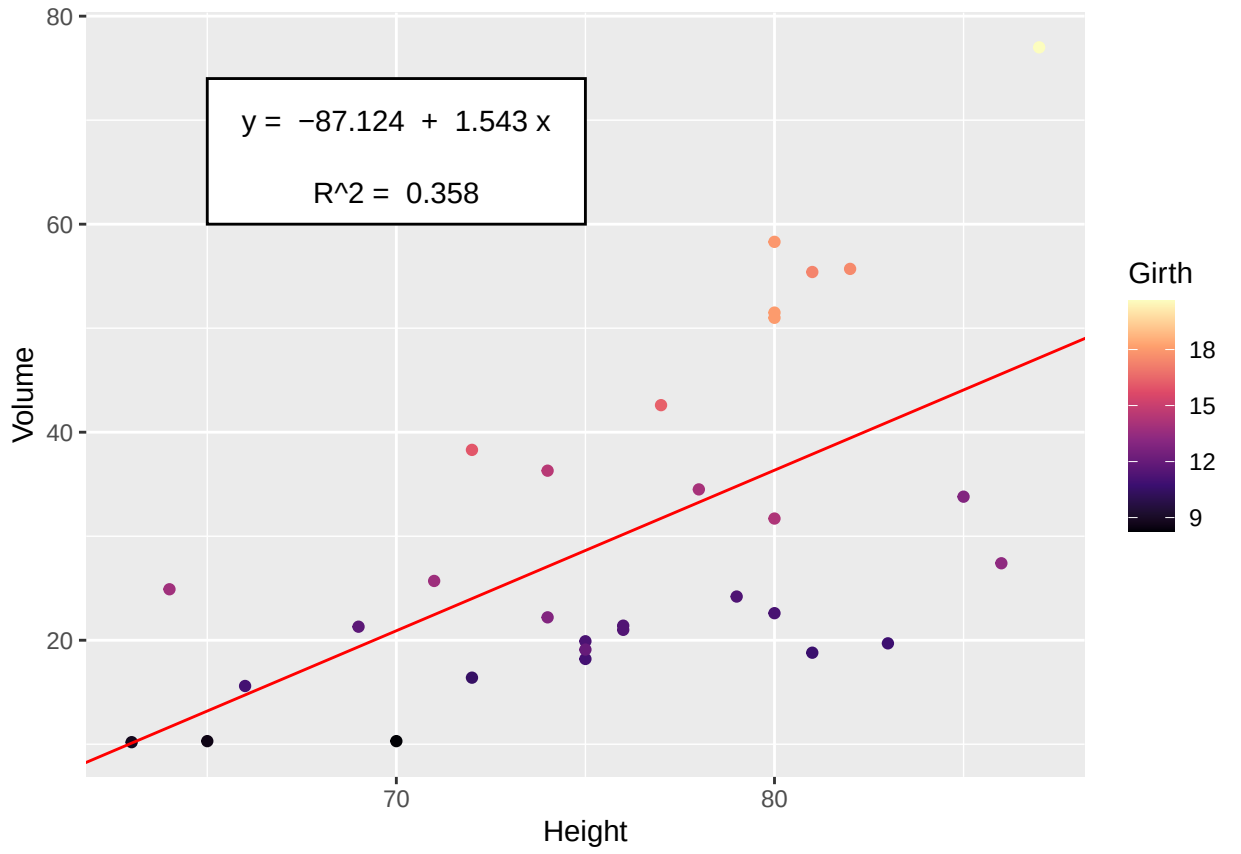
h) Use the `broom` package to extract the coefficient of determination r^2 from the model. Add the annotation to your graph, which should look something like $R^2 = (VALUE)$

```
r.2 <- round(broom::glance(trees.model)$r.squared,3)

ggplot(data = trees, aes(x = Height, y = Volume)) +
  geom_point(aes(color = Girth)) +
  scale_color_viridis_c(option='magma')+
  annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
         fill='white', color='black')+
  annotate("text", x = 70, y = 70,
           label = paste("y = ", int, " + ", slope, "x"), parse = FALSE)+
  annotate("text", x = 70, y = 63,
           label = paste("R^2 = ", r.2), parse = FALSE)
```



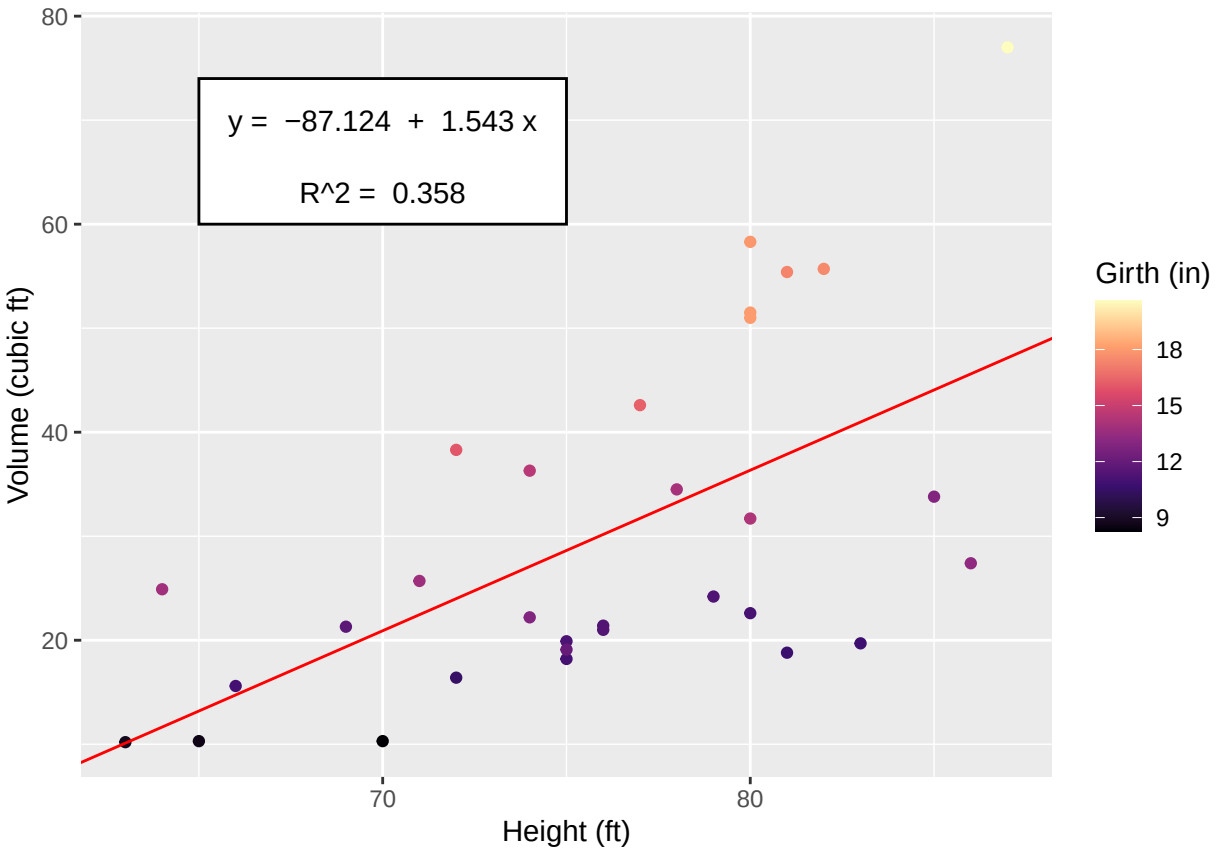
i) Add the regression line in red. There are several ways to do this.

```
SLM.Int <- trees.model$coefficients[1]
SLM.Slope <- trees.model$coefficients[2]

ggplot(data = trees, aes(x = Height, y = Volume)) +
  geom_point(aes(color = Girth)) +
  scale_color_viridis_c(option='magma')+
  annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
         fill='white', color='black')+
  annotate("text", x = 70, y = 70,
           label = paste("y = ", int, " + ", slope, "x"), parse = FALSE)+
  annotate("text", x = 70, y = 63,
           label = paste("R^2 = ", r.2), parse = FALSE)+
  geom_abline(intercept = SLM.Int, slope = SLM.Slope, color = "red")
```



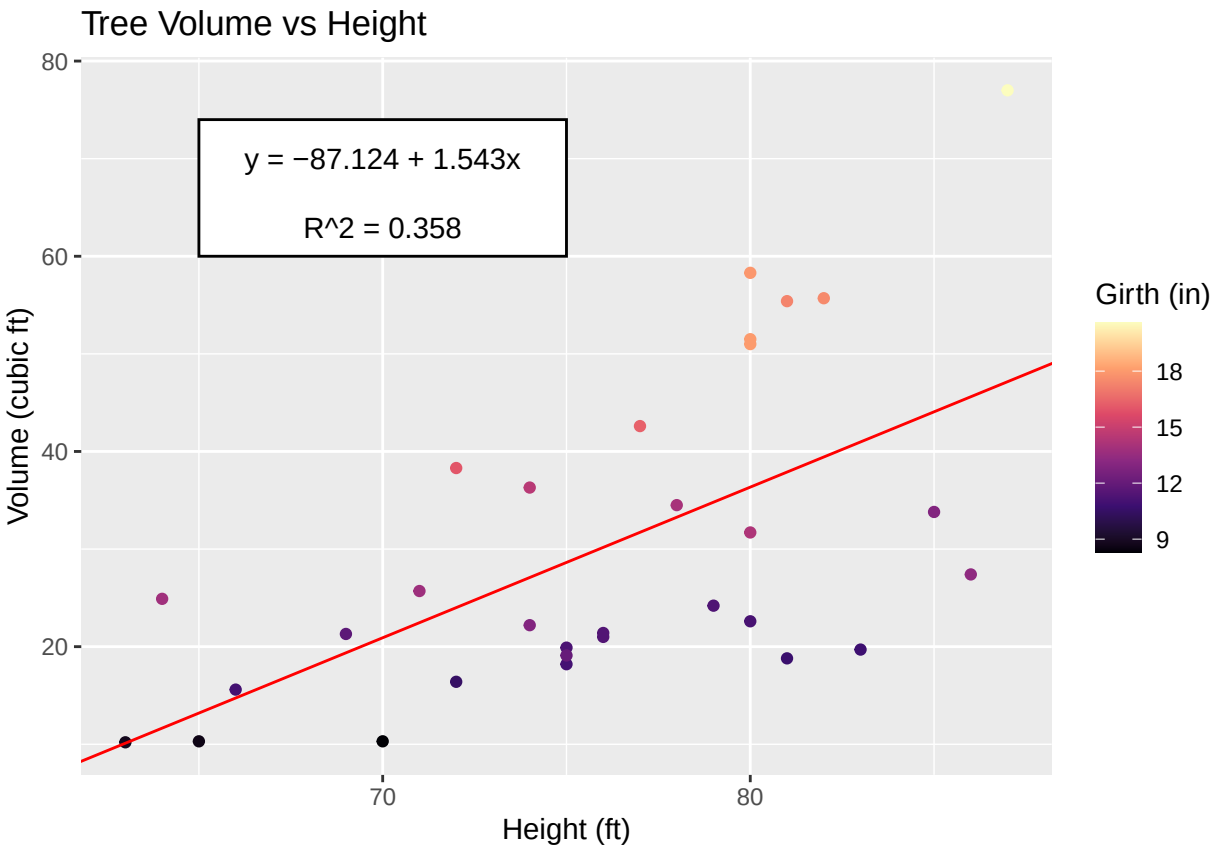
j) Properly label the axes of the graph.

```
ggplot(data = trees, aes(x = Height, y = Volume)) +
  geom_point(aes(color = Girth)) +
  scale_color_viridis_c(option='magma')+
  annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
         fill='white', color='black')+
  annotate("text", x = 70, y = 70,
           label = paste("y = ", int, " + ", slope, "x"), parse = FALSE)+
  annotate("text", x = 70, y = 63,
           label = paste("R^2 = ", r.2), parse = FALSE)+
  geom_abline(intercept = SLM.Int, slope = SLM.Slope, color = "red")+
  labs(x = "Height (ft)",
       y = "Volume (cubic ft)",
       color = "Girth (in)")
```



k) Add a descriptive title to the graph.

```
ggplot(data = trees, aes(x = Height, y = Volume)) +
  geom_point(aes(color = Girth)) +
  scale_color_viridis_c(option='magma')+
  annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
         fill='white', color='black')+
  annotate("text", x = 70, y = 70,
           label = paste0("y = ", int, " + ", slope, "x"), parse = FALSE)+
  annotate("text", x = 70, y = 63,
           label = paste0("R^2 = ", r.2), parse = FALSE)+
  geom_abline(intercept = SLM.Int, slope = SLM.Slope, color = "red")+
  labs(title = "Tree Volume vs Height",
       x = "Height (ft)",
       y = "Volume (cubic ft)",
       color = "Girth (in)")
```


Tree Volume vs Height
y = −87.124 + 1.543x
R^2 = 0.358
Volume (cubic ft)
Height (ft)
80
60
40
20
70
80
Girth (in)
18
15
12
9