Assignment 7

Jack Scherer

2024-11-06

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`.

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

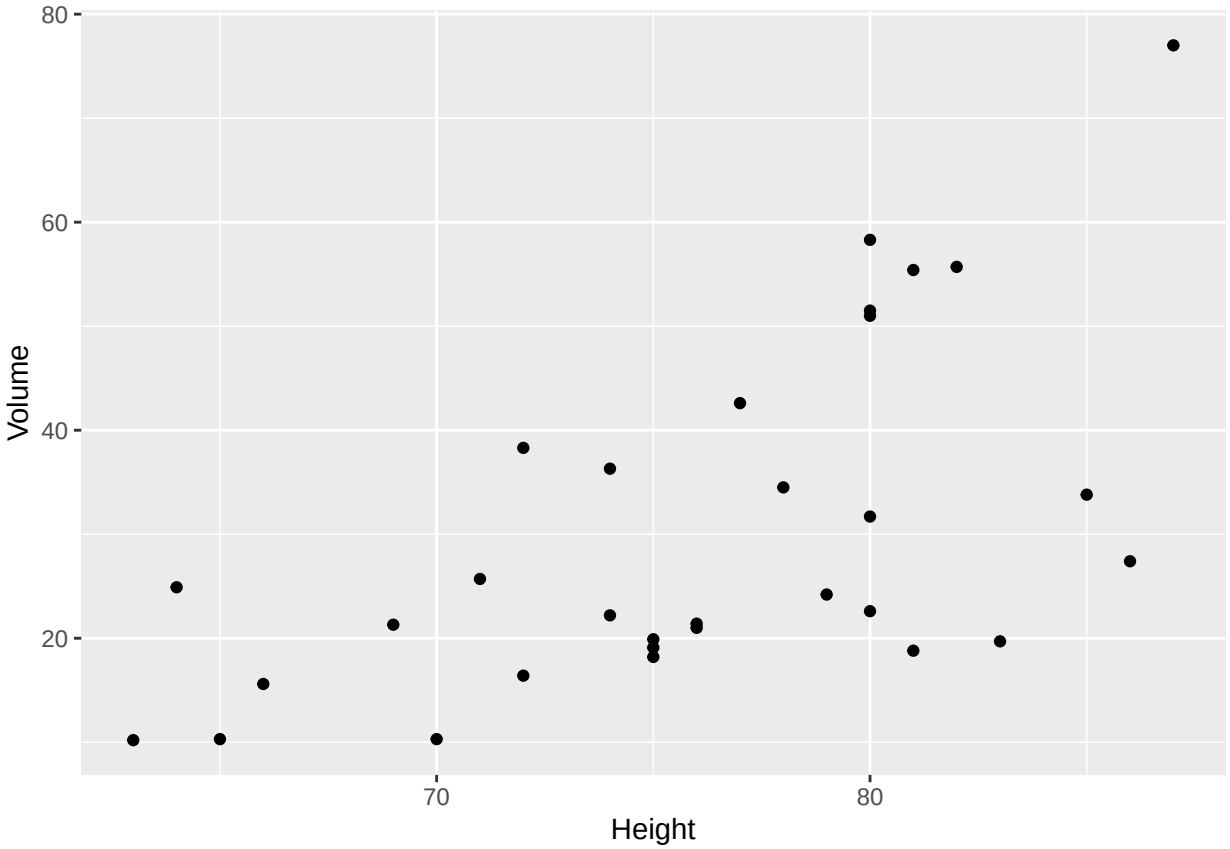
b) Display the `summary` of the model to view the y-intercept and slope of the regression line.

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

```
##
## 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
```

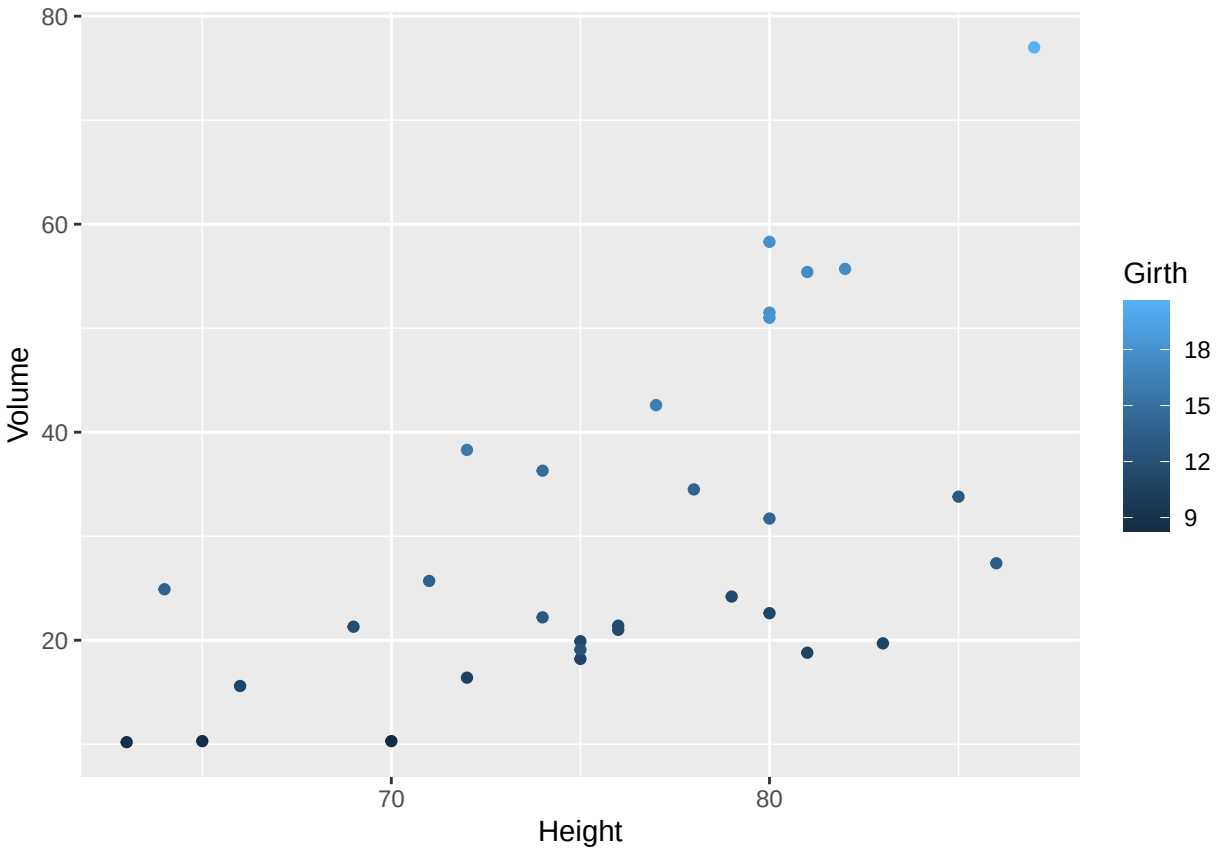
c) Using `ggplot2`, create a scatter plot of Volume vs Height.

```
plot.trees <- ggplot( data=trees, aes(x=Height, y=Volume) ) +
  geom_point()

plot.trees
```



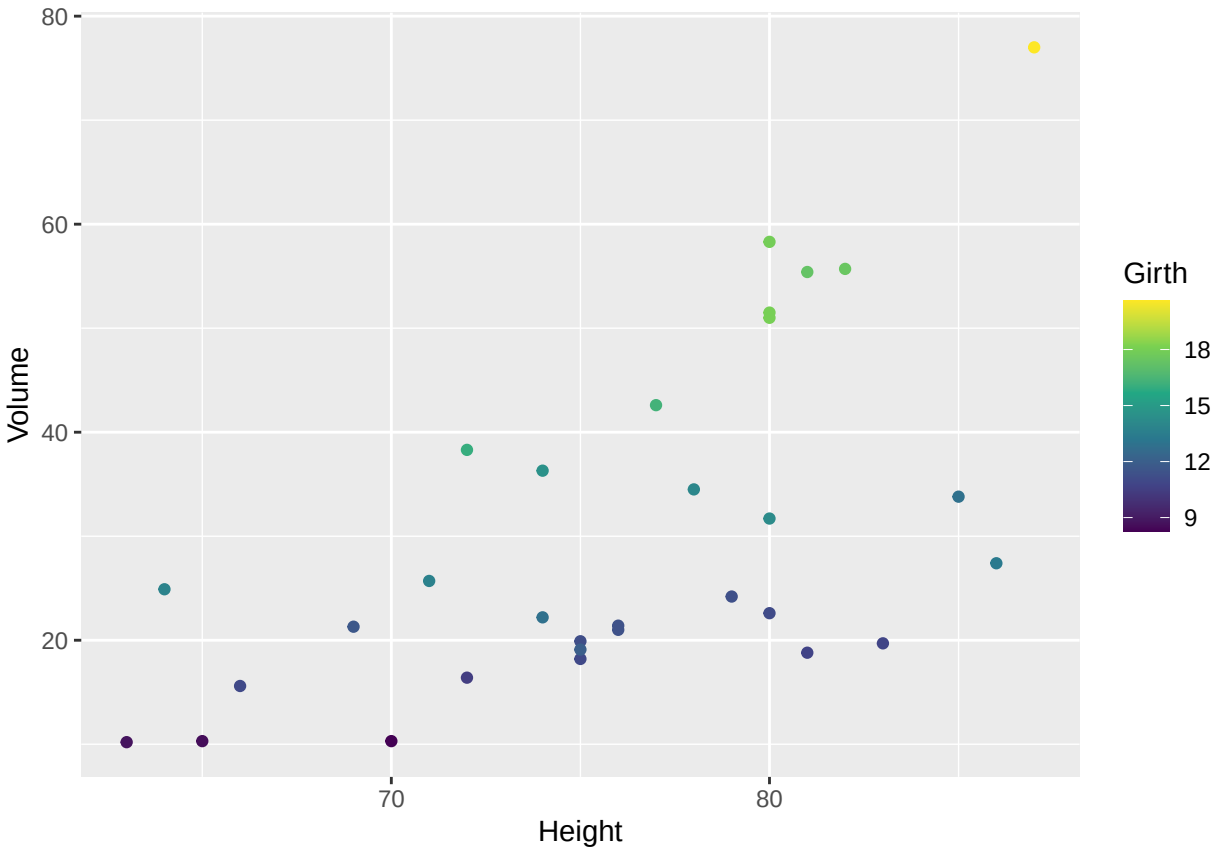
d) Color the scatter using the `Girth` variable.

```
plot.trees.2 <- ggplot( data=trees, aes(x=Height, y=Volume, color=Girth) ) +
  geom_point()

plot.trees.2
```



e) Modify the color scheme using a RColorBrewer palette.

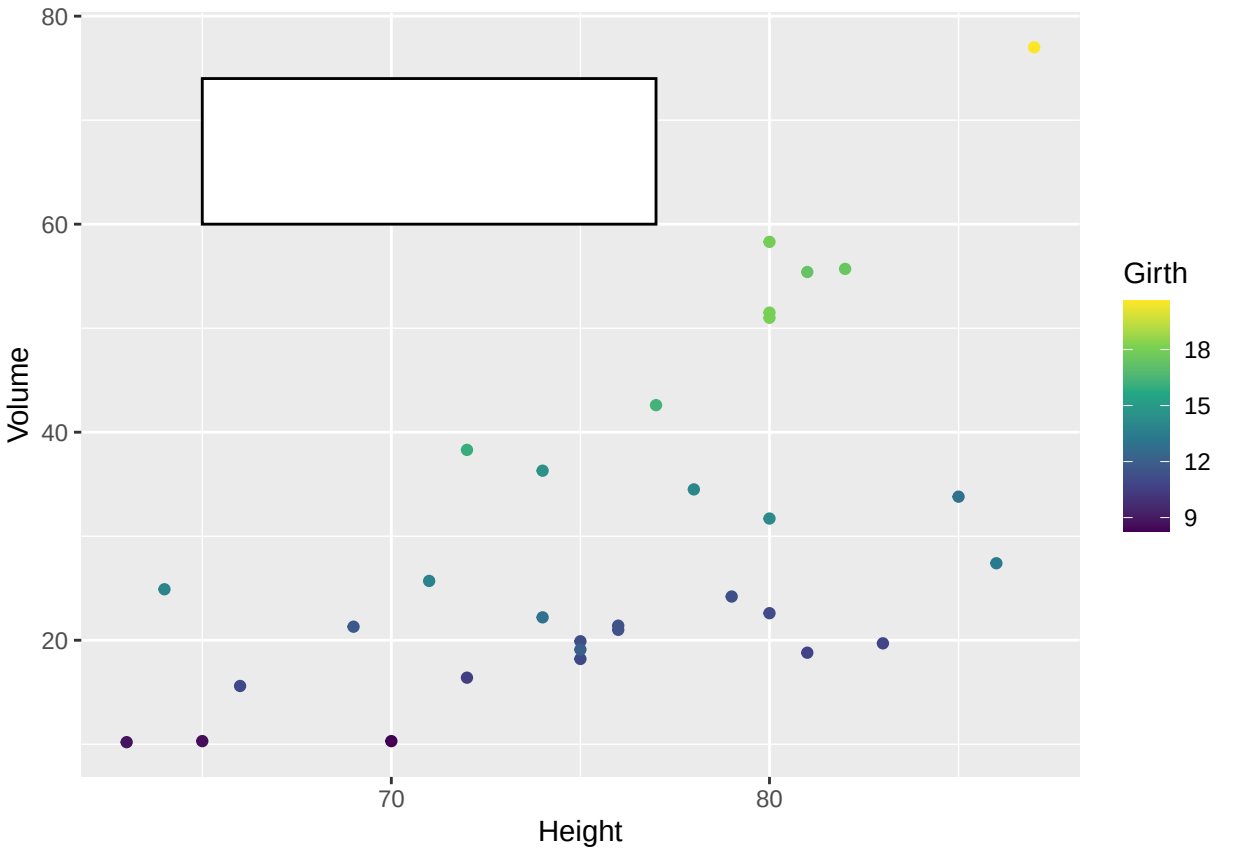
```
plot.trees.3 <- plot.trees.2 +
  #scale_color_distiller( palette = "RdYlBu" ) +
  scale_color_viridis_c(option='viridis')

plot.trees.3
```



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') +
```

```
plot.trees.4 <- plot.trees.3 +
  annotate('rect', xmin=65, xmax=77, ymin=60, ymax=74,
         fill='white', color='black')

plot.trees.4
```

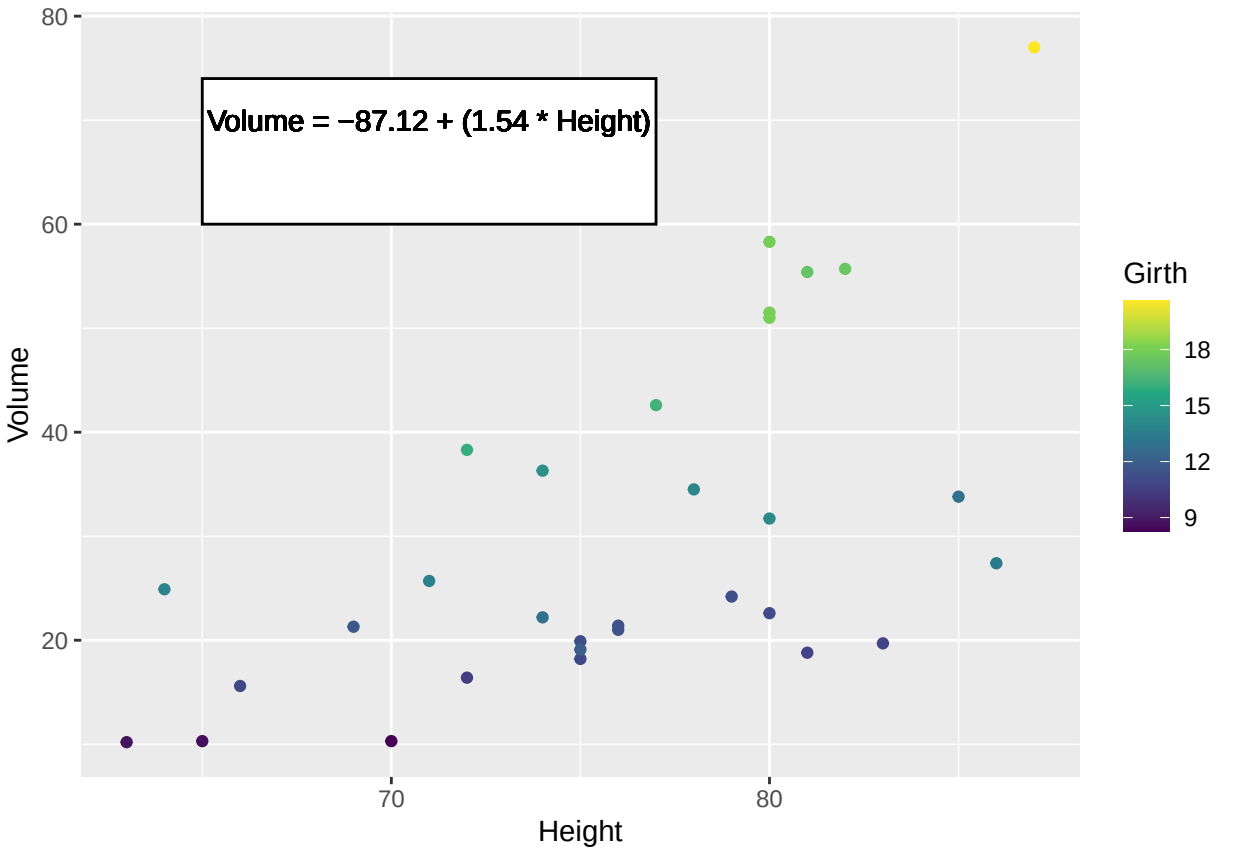


g) Use the `broom` package to extract the coefficients of the best-fit line. Add this information as an annotation to the graph, which should follow a form that looks like $\hat{y}_i = (INTERCEPT) + (SLOPE) * x_i$. Place the annotation within the white text box.

```
broom::tidy( model.trees )
```

```
## # 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
```

```
plot.trees.5 <- plot.trees.4 +
  geom_text( label="Volume = -87.12 + (1.54 * Height)", x=71, y=70, colour='black' )

plot.trees.5
```



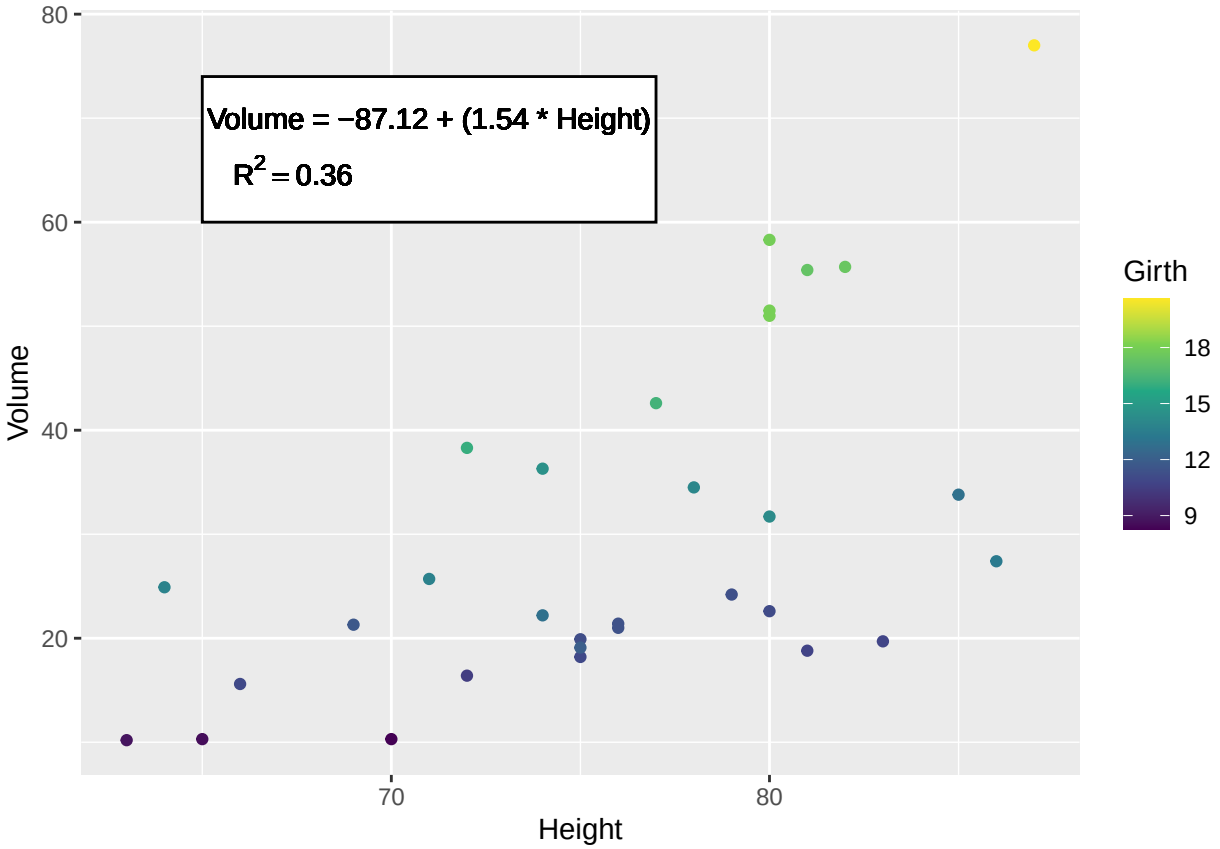
h) Use the `broom` package to extract the coefficient of determination r^2 from the model. Add the annotation to your graph, which should look something like $R^2 = (VALUE)$

```
broom::glance( model.trees )
```

```
## # A tibble: 1 x 12
##   r.squared adj.r.squared sigma statistic  p.value    df logLik   AIC   BIC
##       <dbl>         <dbl> <dbl>     <dbl>    <dbl> <dbl>  <dbl> <dbl> <dbl>
## 1     0.358         0.336  13.4      16.2 0.000378     1  -123.  253.  257.
## # i 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```

```
plot.trees.6 <- plot.trees.5 +
  geom_text( label=latex2exp::TeX('$R^2 = 0.36$'), x=67.4, y=65, colour='black' )

plot.trees.6
```

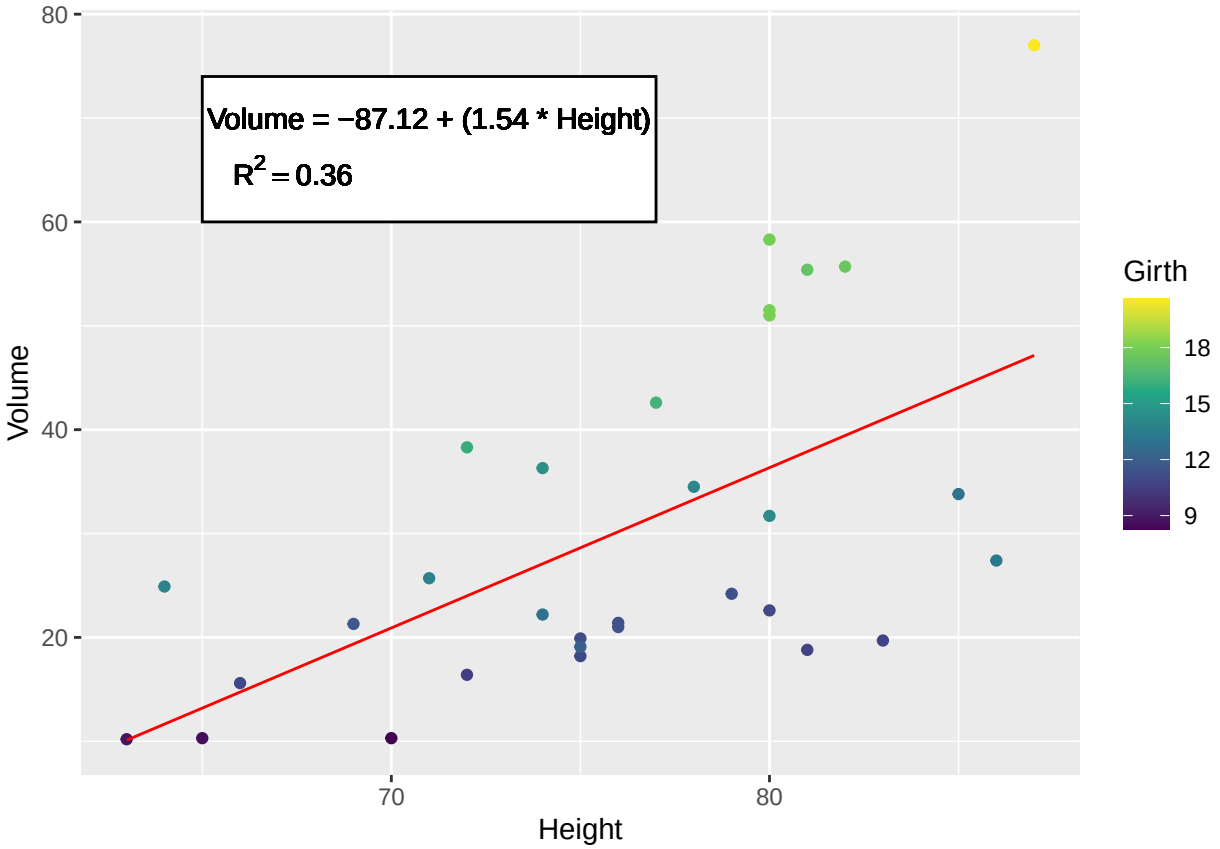
```
## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'
```



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

```
trees <- trees %>%
  mutate( fit = fitted(model.trees) )

plot.trees.7 <- plot.trees.6 +
  geom_line( aes( y=trees$fit ), colour='red' )

plot.trees.7
```

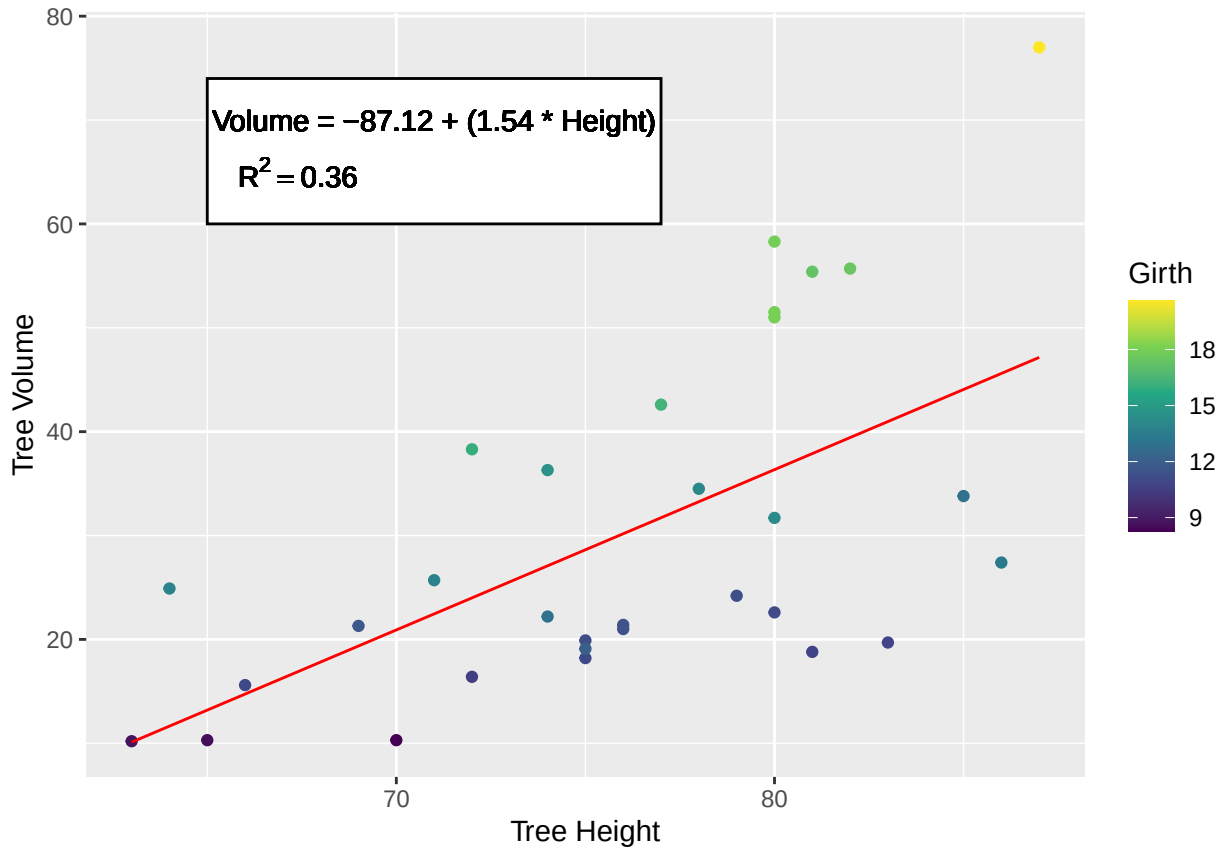
```
## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'
```



j) Properly label the axes of the graph.

```
plot.trees.8 <- plot.trees.7 +
  labs( x='Tree Height', y='Tree Volume' )

plot.trees.8
```

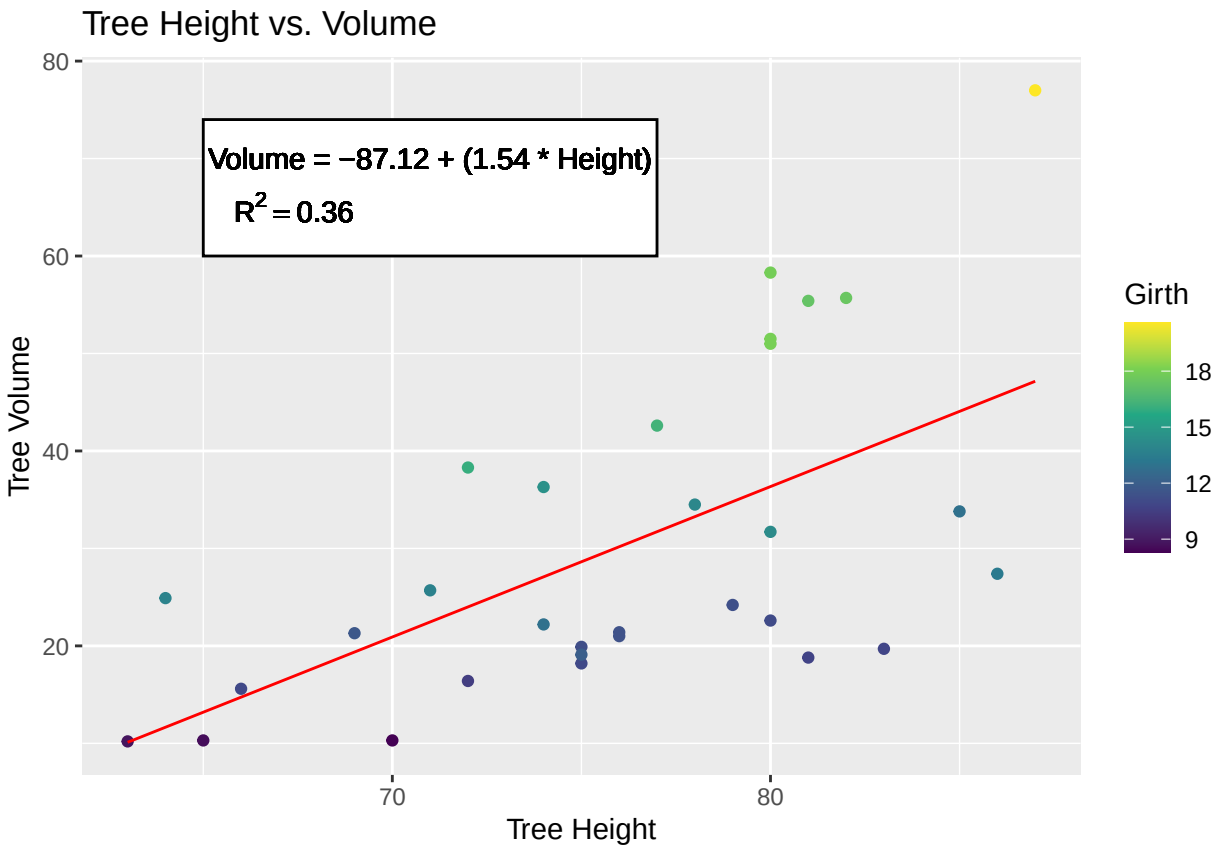
```
## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'
```



k) Add a descriptive title to the graph.

```
plot.trees.9 <- plot.trees.8 +
  labs( title='Tree Height vs. Volume' )

plot.trees.9
```

```
## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'
```


Tree Height vs. Volume
Volume = −87.12 + (1.54 * Height)
$R^2 = 0.36$
Tree Volume
Tree Height
80
60
40
20
70
80
Girth
18
15
12
9