STA 445 Assignment #7

Chip Haskins

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`.

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

```
##
## Call:
## lm(formula = Volume ~ Height, data = trees)
##
## Coefficients:
## (Intercept)       Height
##     -87.124        1.543
```

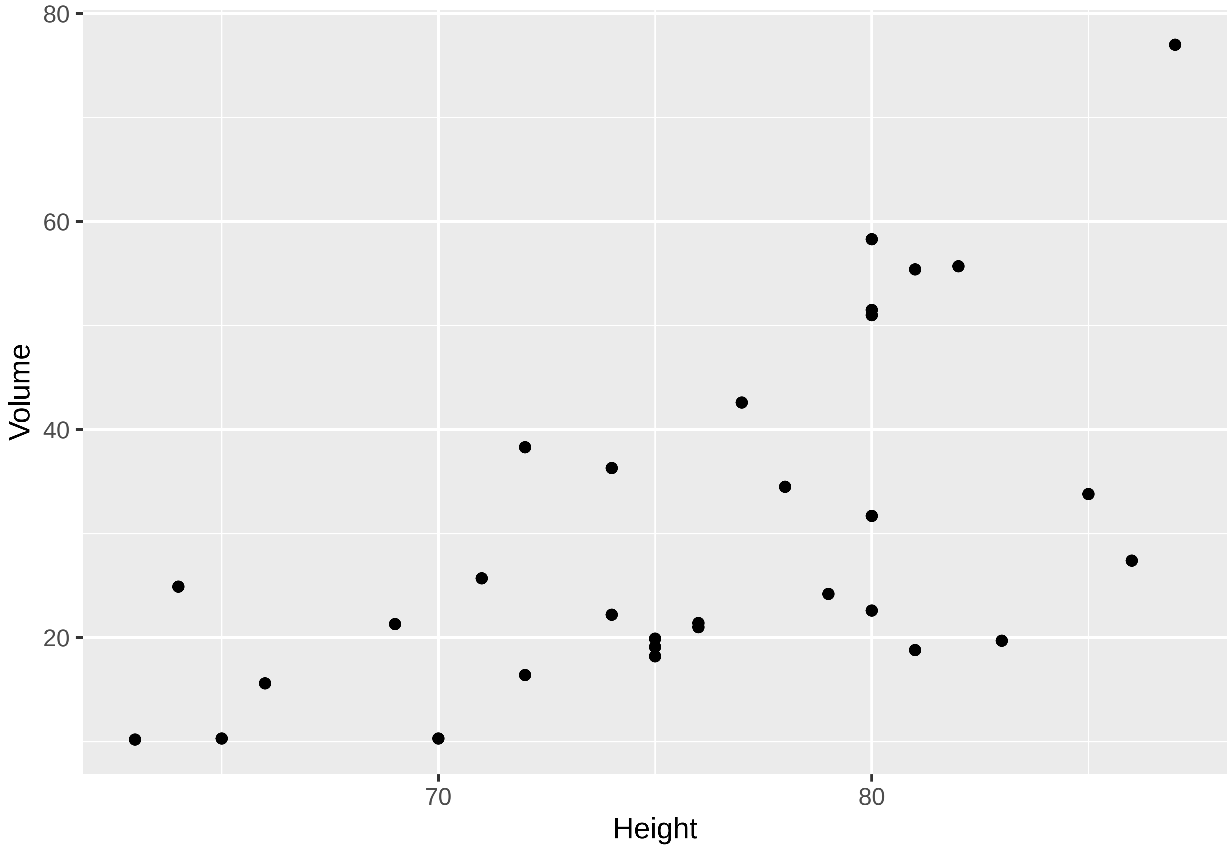
b) Display the `summary` of the model to view the y-intercept and slope of the regression line.

```
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
```

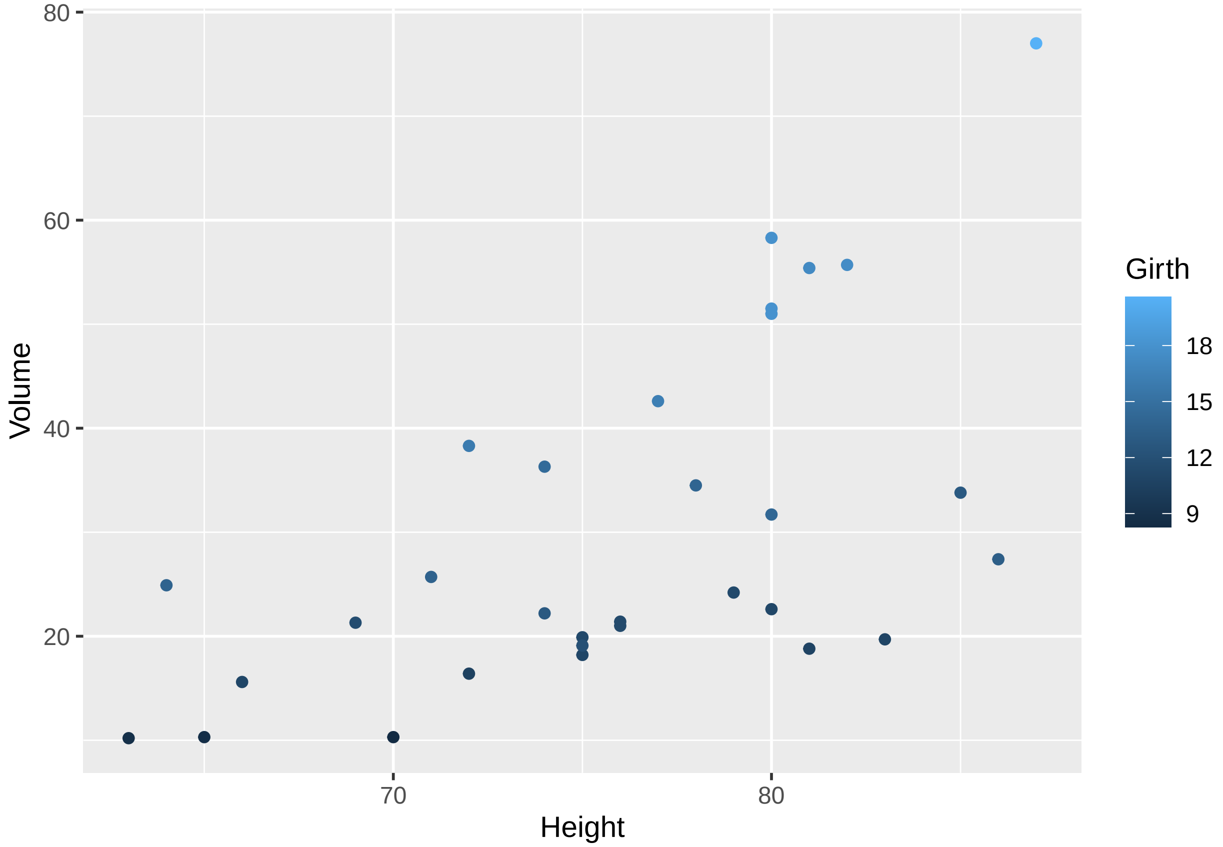
c) Using ggplot2, create a scatter plot of Volume vs Height.

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



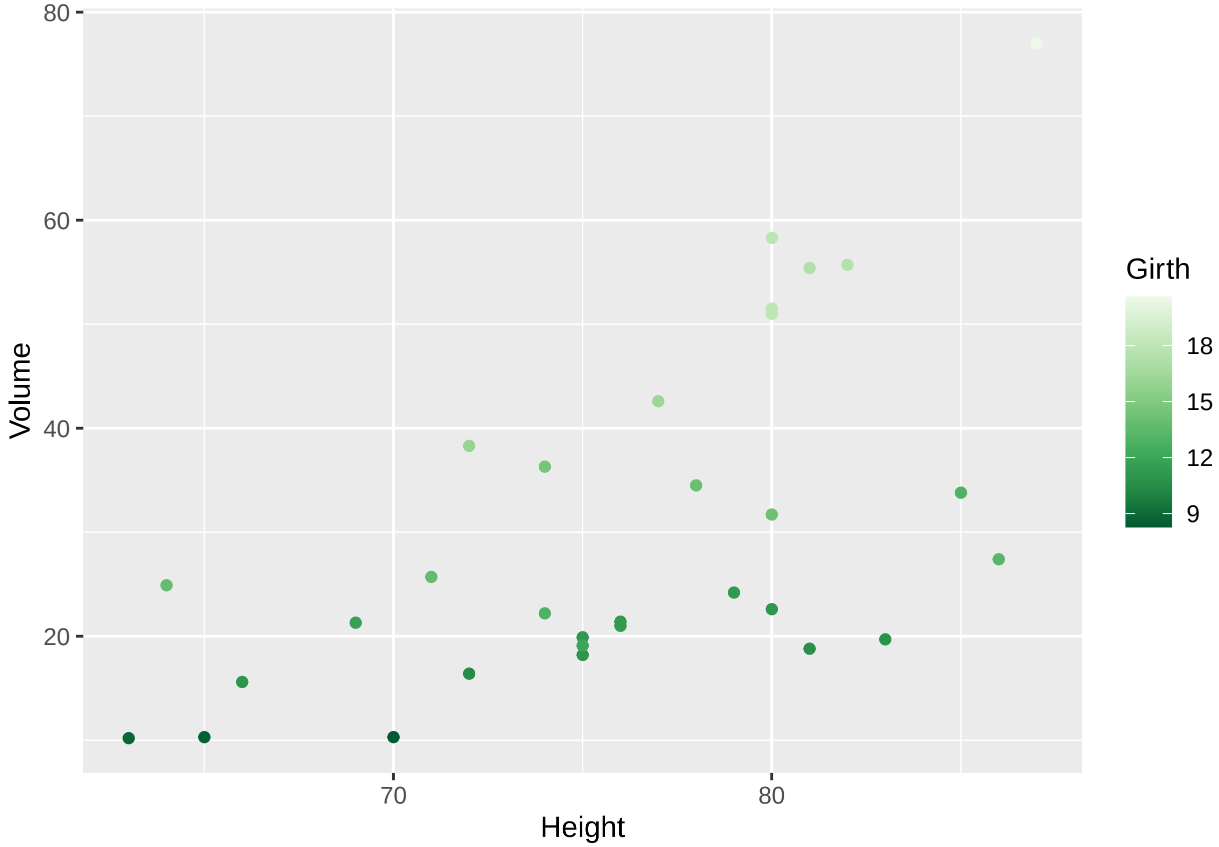
d) Color the scatter using the Girth variable.

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



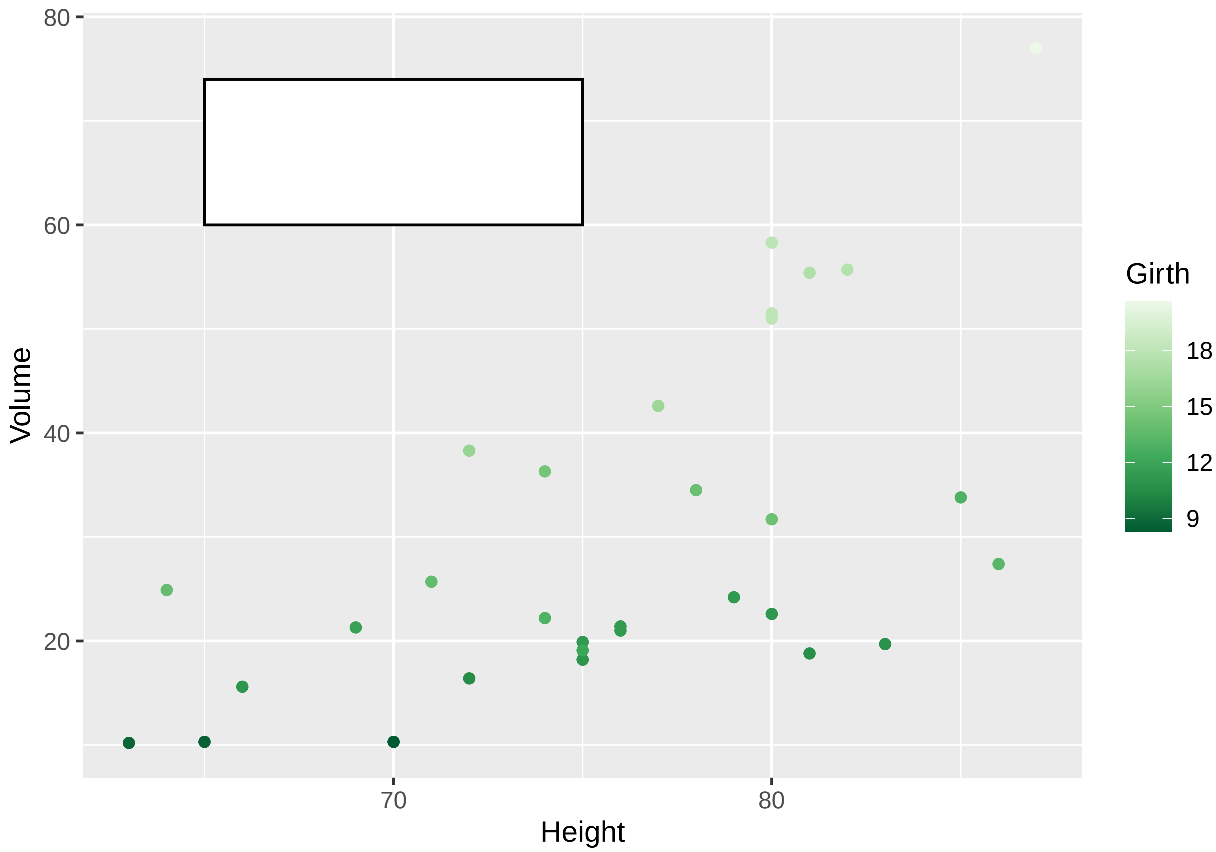
e) Modify the color scheme using a RColorBrewer palette.

```
ggplot(trees, aes(x = Height, y = Volume)) +
  geom_point(aes(color = Girth)) +
  scale_color_distiller(palette = "Greens")
```



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

```
ggplot(trees, aes(x = Height, y = Volume)) +
  geom_point(aes(color = Girth)) +
  scale_color_distiller(palette = "Greens") +
  annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
         fill='white', color='black')
```

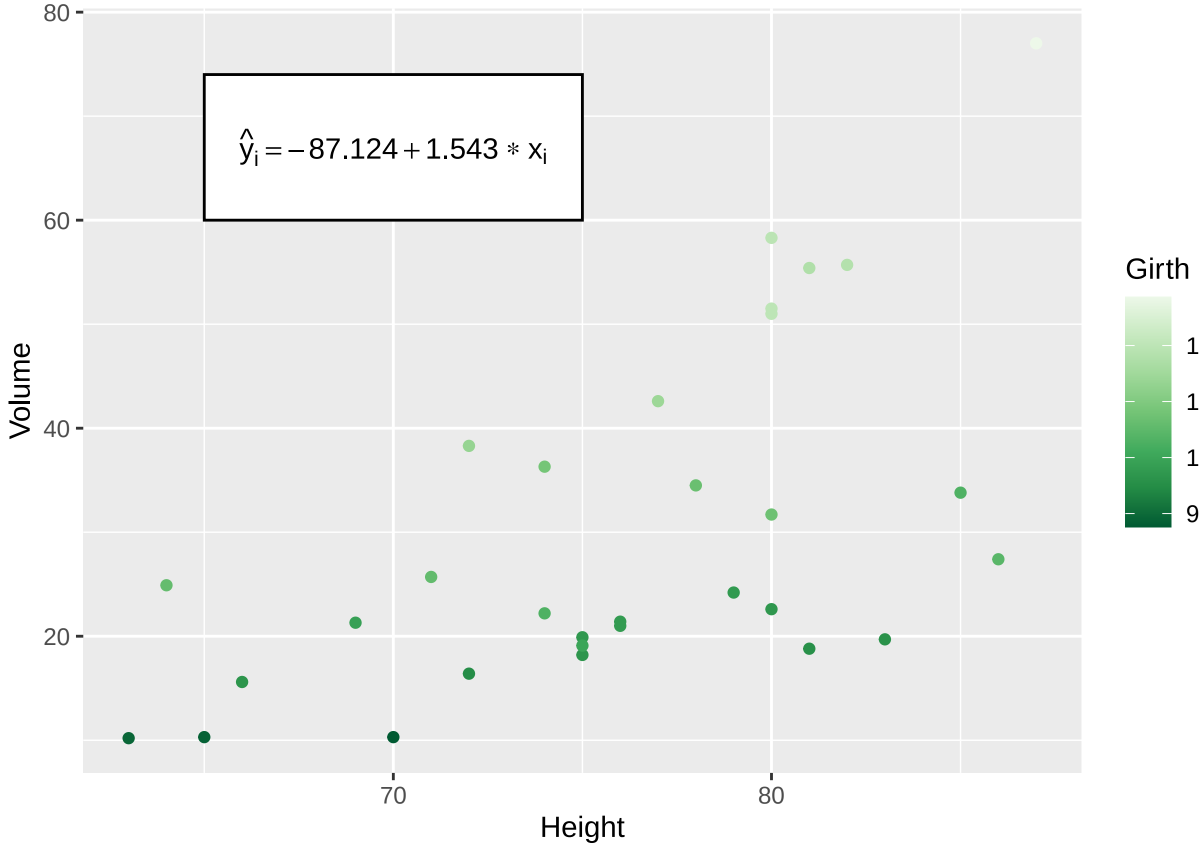


g) Use the broom package to extract the coefficients of the best-fit line. Add this information as an annotation to the graph, which should follow a form that looks like $\hat{y}_i = (INTERCEPT) + (SLOPE) * x_i$. Place the annotation within the white text box.

```
trees.lm.coefs <- broom::tidy(trees.lm)
trees.lm.intercept <- filter(trees.lm.coefs, term == "(Intercept)") %>%
  pull(estimate) %>%
  round(3)
trees.lm.slope <- filter(trees.lm.coefs, term == "Height") %>%
  pull(estimate) %>%
  round(3)

ggplot(trees, aes(x = Height, y = Volume)) +
  geom_point(aes(color = Girth)) +
  scale_color_distiller(palette = "Greens") +
  annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
         fill='white', color='black') +
  annotate("text", x = 70, y = 67,
           label = TeX(r"($\hat{y}_i = \intercept + \slope*x_i$)",
                       user_defined = list("\\intercept" = trees.lm.intercept,
                                           "\\slope" = trees.lm.slope)))
```

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

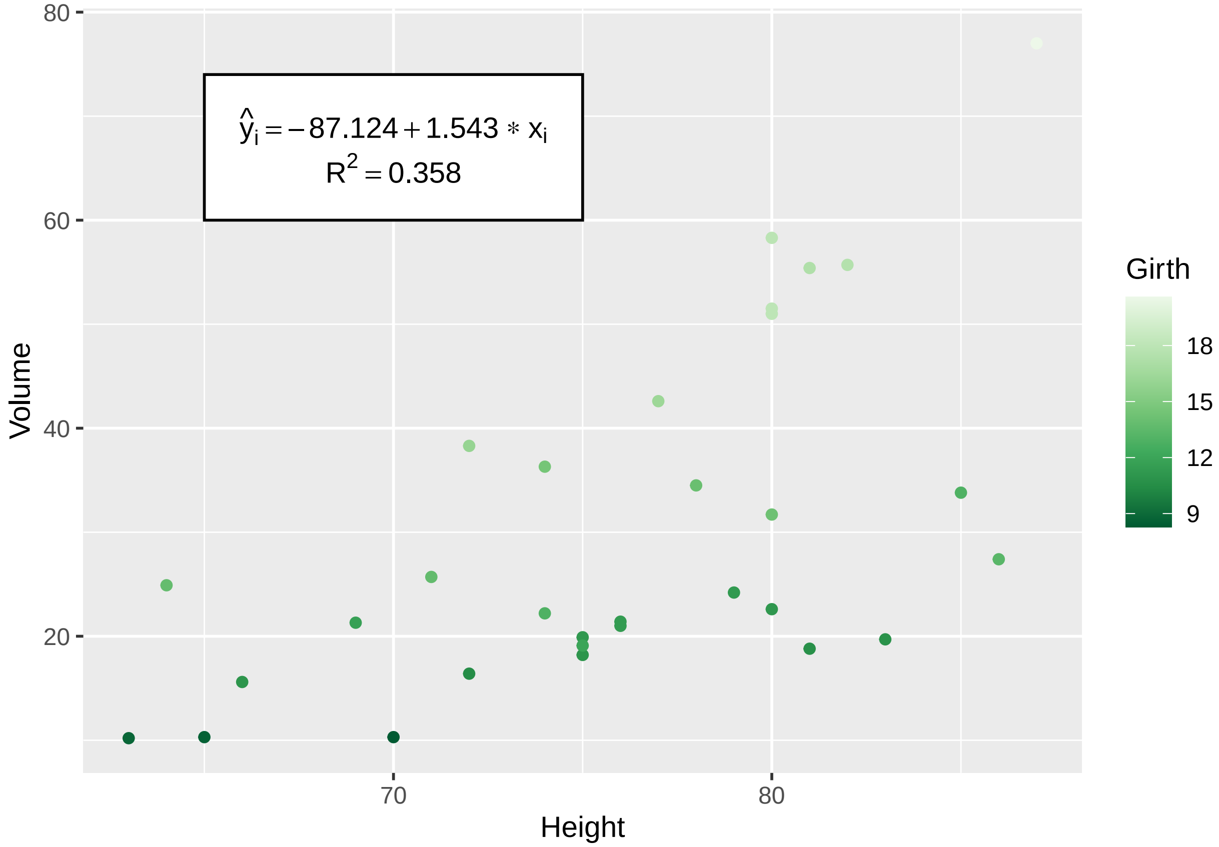


h) Use the broom package to extract the coefficient of determination r^2 from the model. Add the annotation to your graph, which should look something like $R^2 = (VALUE)$

```
trees.lm.rsquared = broom::glance(trees.lm) %>%
  pull(r.squared) %>%
  round(3)

ggplot(trees, aes(x = Height, y = Volume)) +
  geom_point(aes(color = Girth)) +
  scale_color_distiller(palette = "Greens") +
  annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
         fill='white', color='black') +
  annotate("text", x = 70, y = 69,
          label = TeX(r"($\hat{y}_i = \intercept + \slope*x_i$)",
                      user_defined = list("\\intercept" = trees.lm.intercept,
                                          "\\slope" = trees.lm.slope))) +
  annotate("text", x = 70, y = 65,
          label = TeX(r"($R^2 = \rsquared$)",
                      user_defined = list("\\rsquared" = trees.lm.rsquared)))
```

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

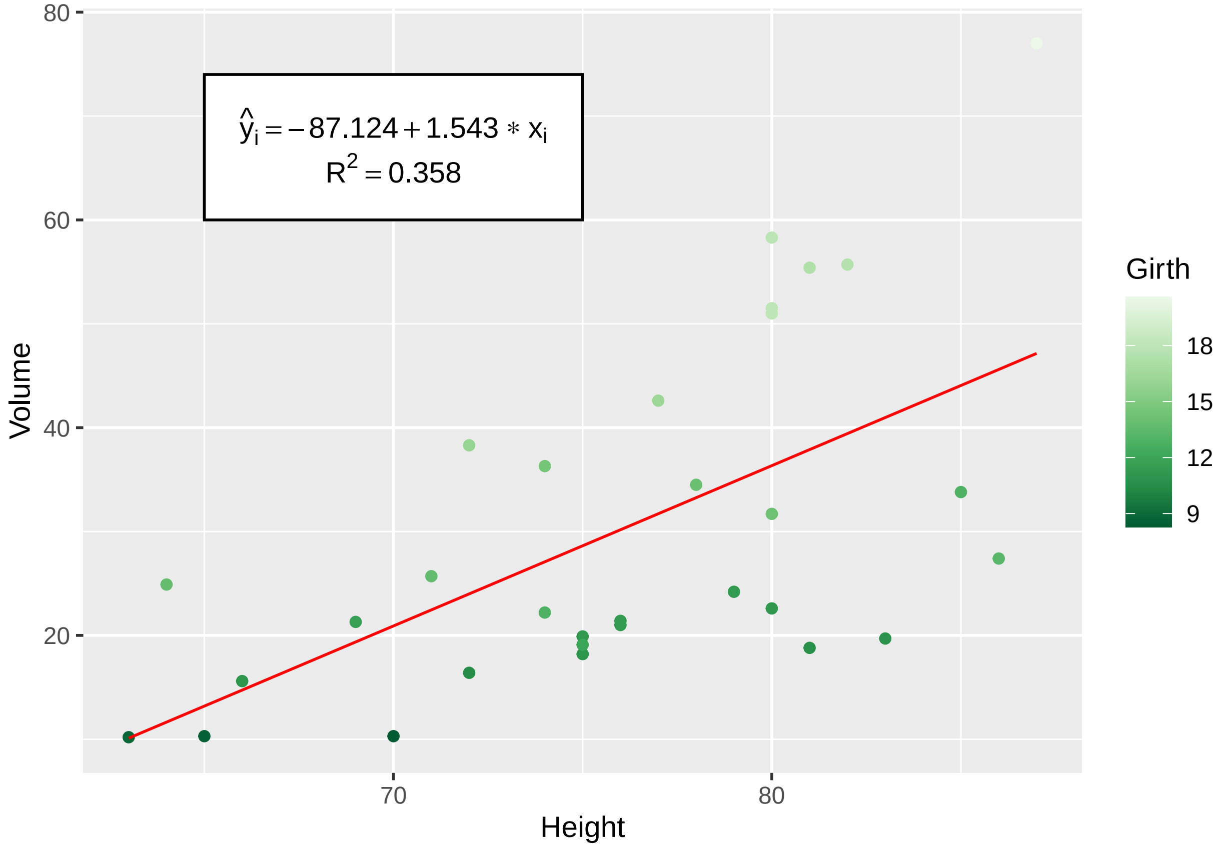


i) Add the regression line in red. There are several ways to do this.

```
trees.predicted <- broom::augment(trees.lm, trees)

ggplot(trees, aes(x = Height, y = Volume)) +
  geom_point(aes(color = Girth)) +
  scale_color_distiller(palette = "Greens") +
  annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
         fill='white', color='black') +
  annotate("text", x = 70, y = 69,
           label = TeX(r"($\hat{y}_i = \intercept + \slope*x_i$)",
                       user_defined = list("\\intercept" = trees.lm.intercept,
                                           "\\slope" = trees.lm.slope))) +
  annotate("text", x = 70, y = 65,
           label = TeX(r"($R^2 = \rsquared$)",
                       user_defined = list("\\rsquared" = trees.lm.rsquared))) +
  geom_line(data = trees.predicted, aes(y = .fitted), color = "red")
```

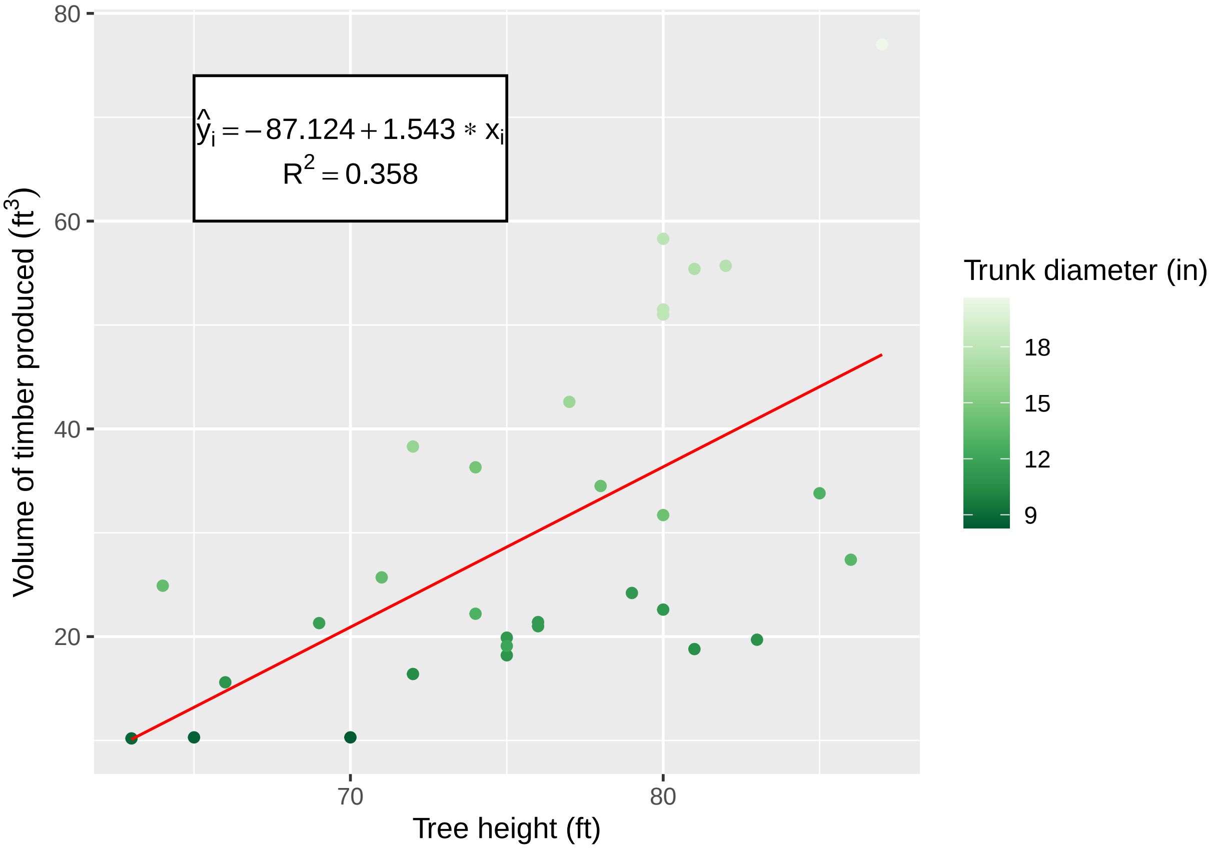
```
## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'
## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'
```



j) Properly label the axes of the graph.

```
ggplot(trees, aes(x = Height, y = Volume)) +
  geom_point(aes(color = Girth)) +
  scale_color_distiller(palette = "Greens") +
  annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
         fill='white', color='black') +
  annotate(
    "text",
    x = 70,
    y = 69,
    label = TeX(
      r"($\hat{y}_i = \intercept + \slope*x_i$)",
      user_defined = list("\\intercept" = trees.lm.intercept,
                          "\\slope" = trees.lm.slope)
      )
    ) +
  annotate(
    "text",
    x = 70,
    y = 65,
    label = TeX(
      r"($R^2 = \rsquared$)",
      user_defined = list("\\rsquared" = trees.lm.rsquared)
      )
    ) +
  geom_line(data = trees.predicted, aes(y = .fitted), color = "red") +
  labs(
    x = "Tree height (ft)",
    y = TeX("Volume of timber produced $(ft^3)$"),
    color = "Trunk diameter (in)"
    )
```

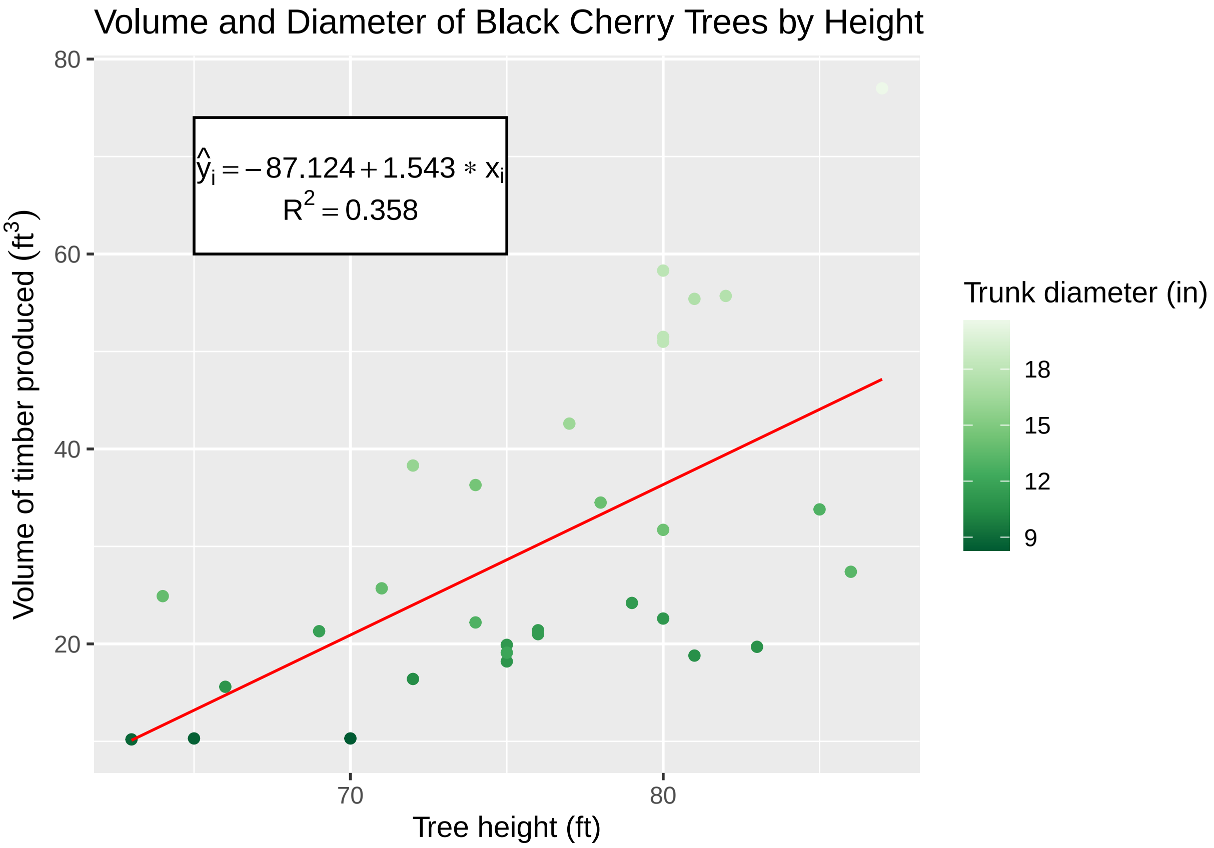
```
## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'
## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'
```



k) Add a descriptive title to the graph.

```
ggplot(trees, aes(x = Height, y = Volume)) +
  geom_point(aes(color = Girth)) +
  scale_color_distiller(palette = "Greens") +
  annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74,
           fill='white', color='black') +
  annotate(
    "text",
    x = 70,
    y = 69,
    label = TeX(
      r"($\hat{y}_i = \intercept + \slope*x_i$)",
      user_defined = list("\\intercept" = trees.lm.intercept,
                          "\\slope" = trees.lm.slope)
      )
    ) +
  annotate(
    "text",
    x = 70,
    y = 65,
    label = TeX(
      r"($R^2 = \rsquared$)",
      user_defined = list("\\rsquared" = trees.lm.rsquared)
      )
    ) +
  geom_line(data = trees.predicted, aes(y = .fitted), color = "red") +
  labs(
    title = "Volume and Diameter of Black Cherry Trees by Height",
    x = "Tree height (ft)",
    y = TeX("Volume of timber produced $(ft^3)$"),
    color = "Trunk diameter (in)"
    )
```

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


Volume and Diameter of Black Cherry Trees by Height
$\hat{y}_i = -87.124 + 1.543 * x_i$
$R^2 = 0.358$
Volume of timber produced (ft^3)
Tree height (ft)
Trunk diameter (in)