Assignment 7

Reagan Anderson

2024-11-09

##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)
```

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

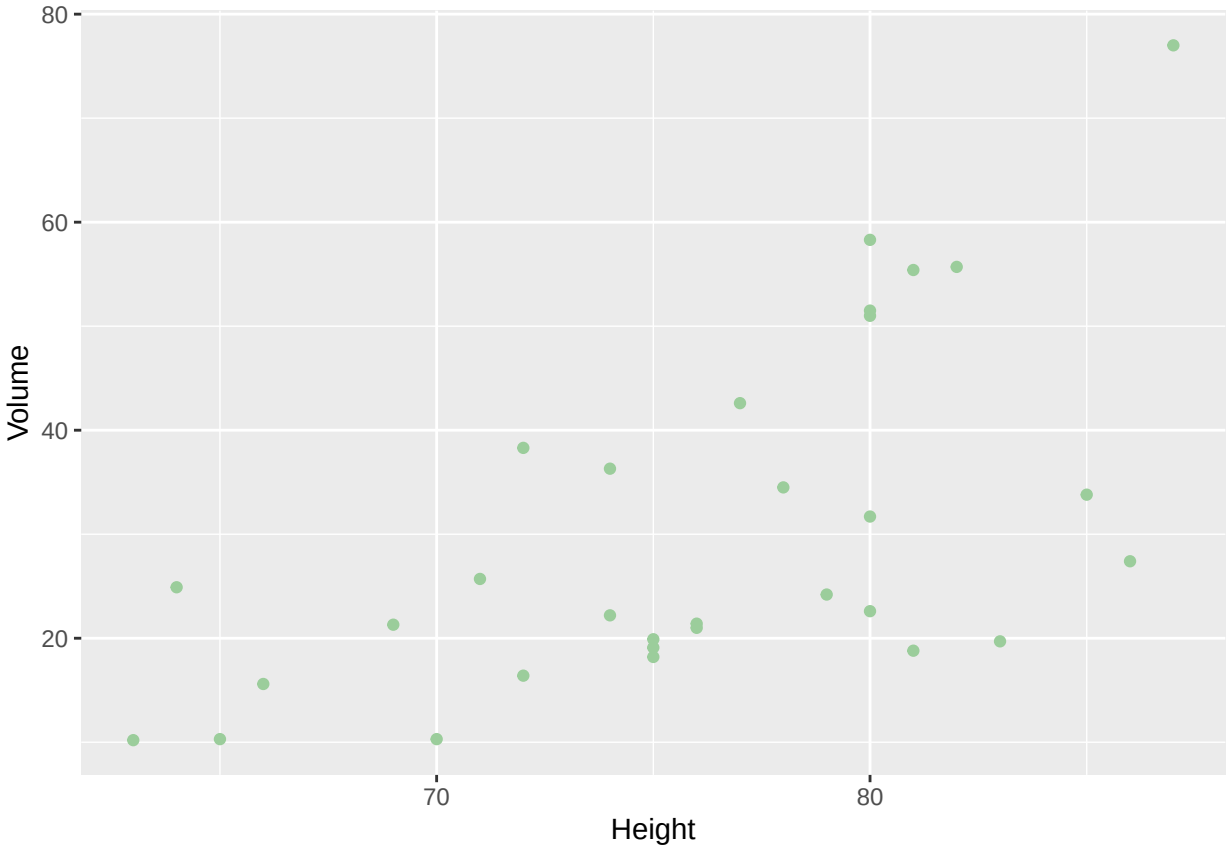
b) Display the `summary` of the model to view the y-intercept and slope of the regression line.

```
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
```

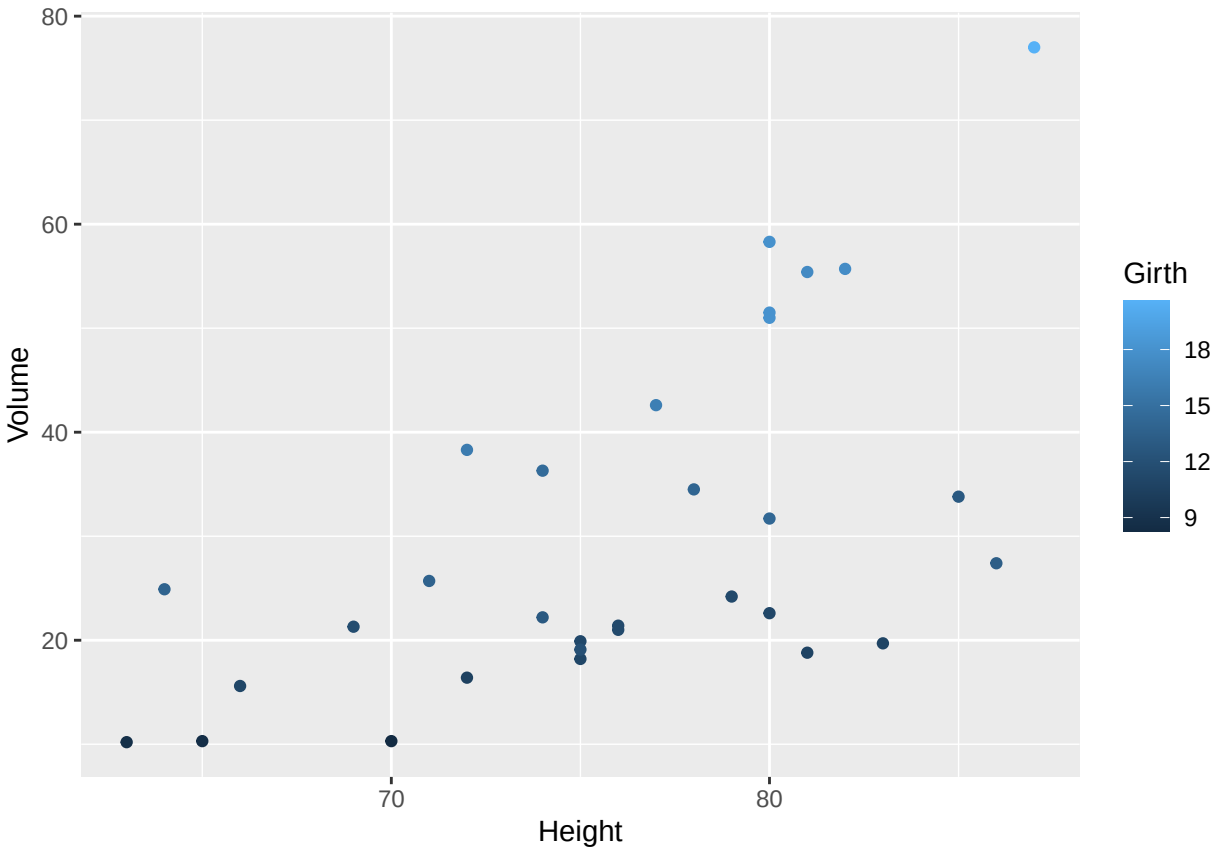
c) Using `ggplot2`, create a scatter plot of Volume vs Height.

```
ggplot(trees,
       aes(x=Height, y = Volume))+
  geom_point(color="darkseagreen3")
```



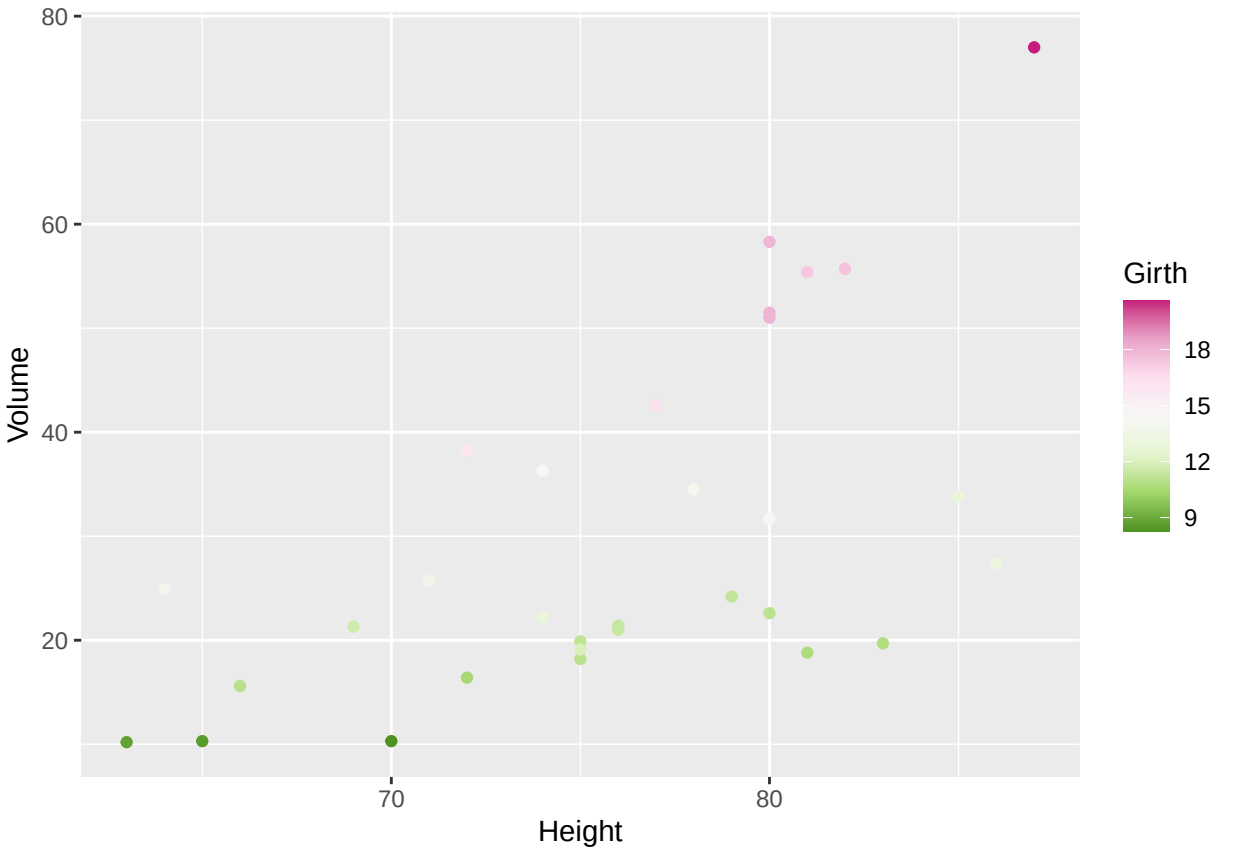
d) Color the scatter using the `Girth` variable.

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



e) Modify the color scheme using a RColorBrewer palette.

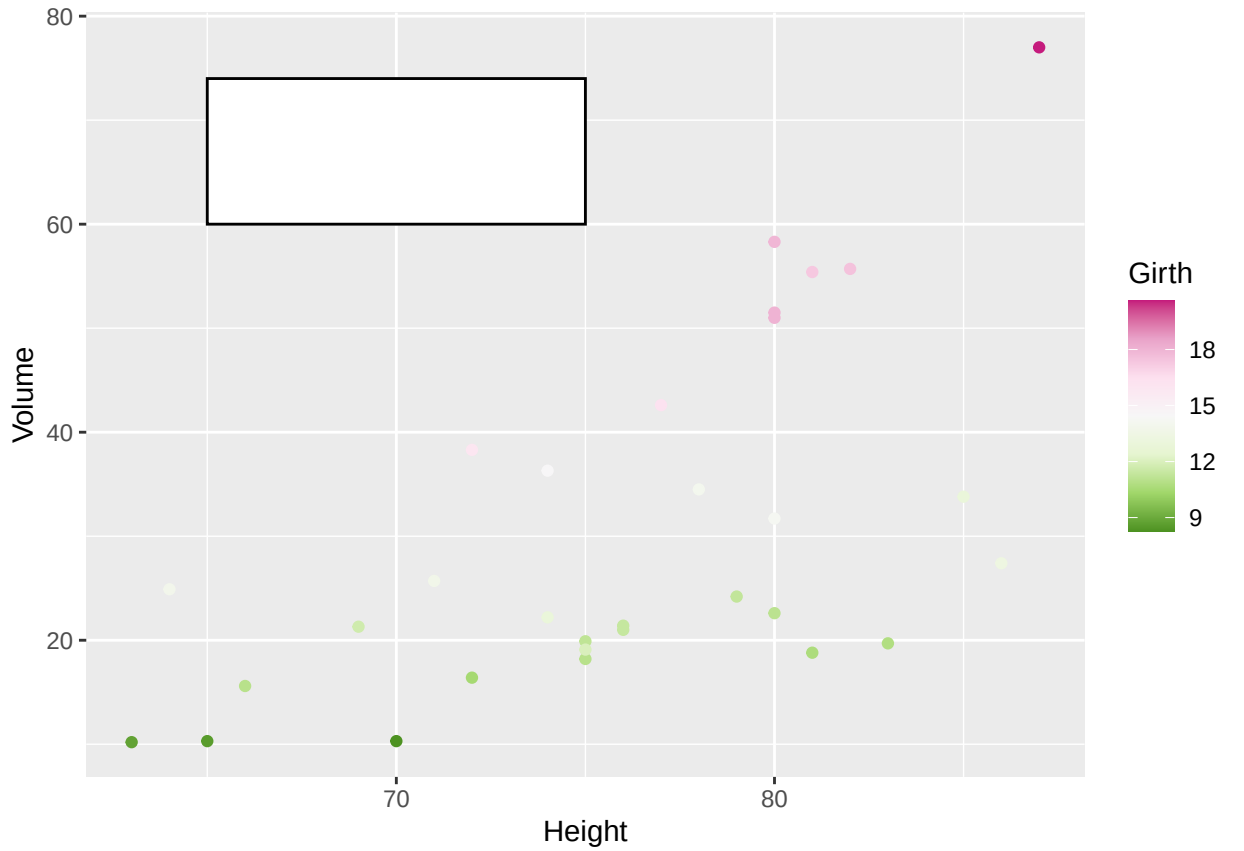
```
ggplot(trees,
       aes(x=Height, y = Volume))+
  geom_point(aes(color=Girth)) +
  scale_color_distiller(palette = 'PiYG')
```



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



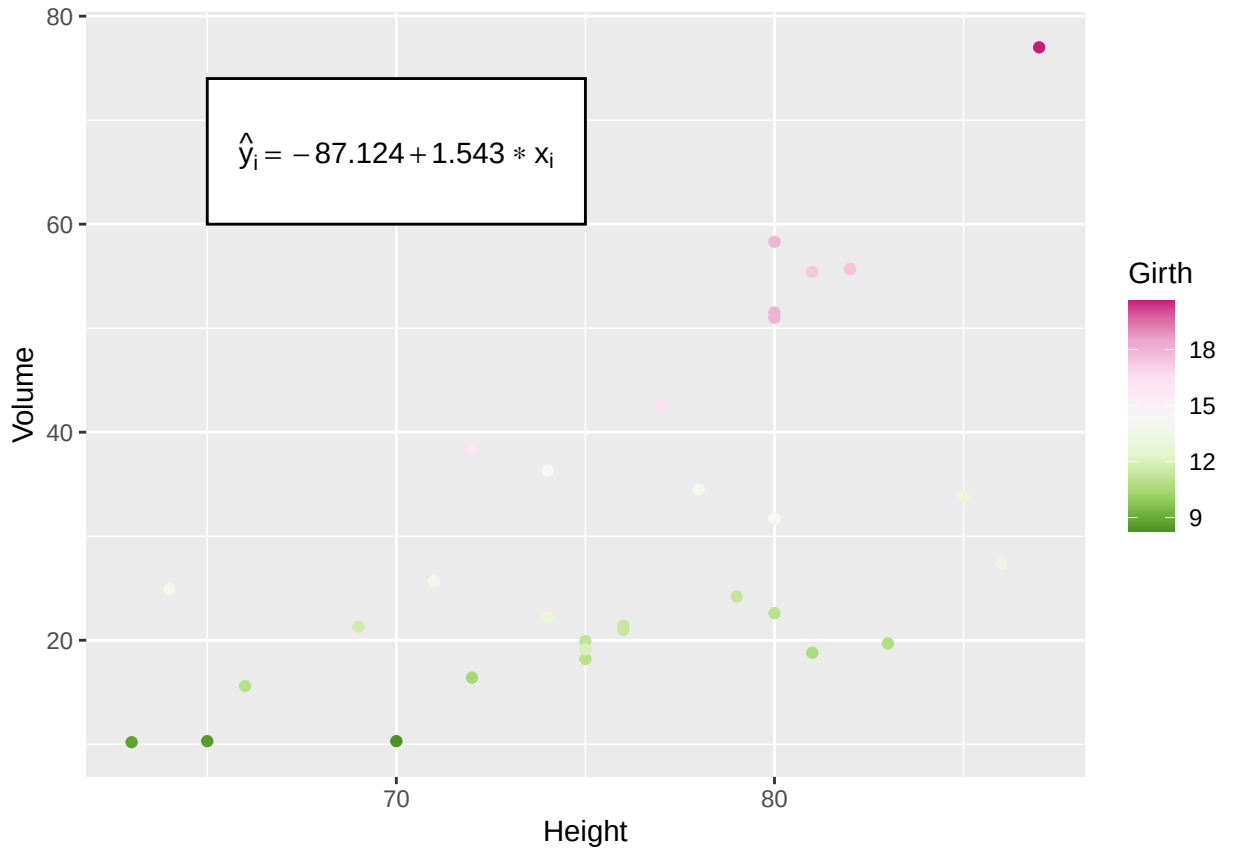
g) Use the `broom` package to extract the coefficients of the best-fit line. Add this information as an annotation to the graph, which should follow a form that looks like $\hat{y}_i = (INTERCEPT) + (SLOPE) * x_i$. Place the annotation within the white text box.

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

```
## [1] -87.12361   1.54335
```

```
ggplot(trees,
       aes(x=Height, y = Volume)) +
  geom_point(aes(color=Girth)) +
  scale_color_distiller(palette = 'PiYG') +
  annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74, fill='white', color='black') +
  annotate('text', x = 70, y = 67, label=latex2exp::TeX('$\\hat{y}_i = -87.124 + 1.543* x_i$'))
```

```
## 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 $R^2 = (VALUE)$

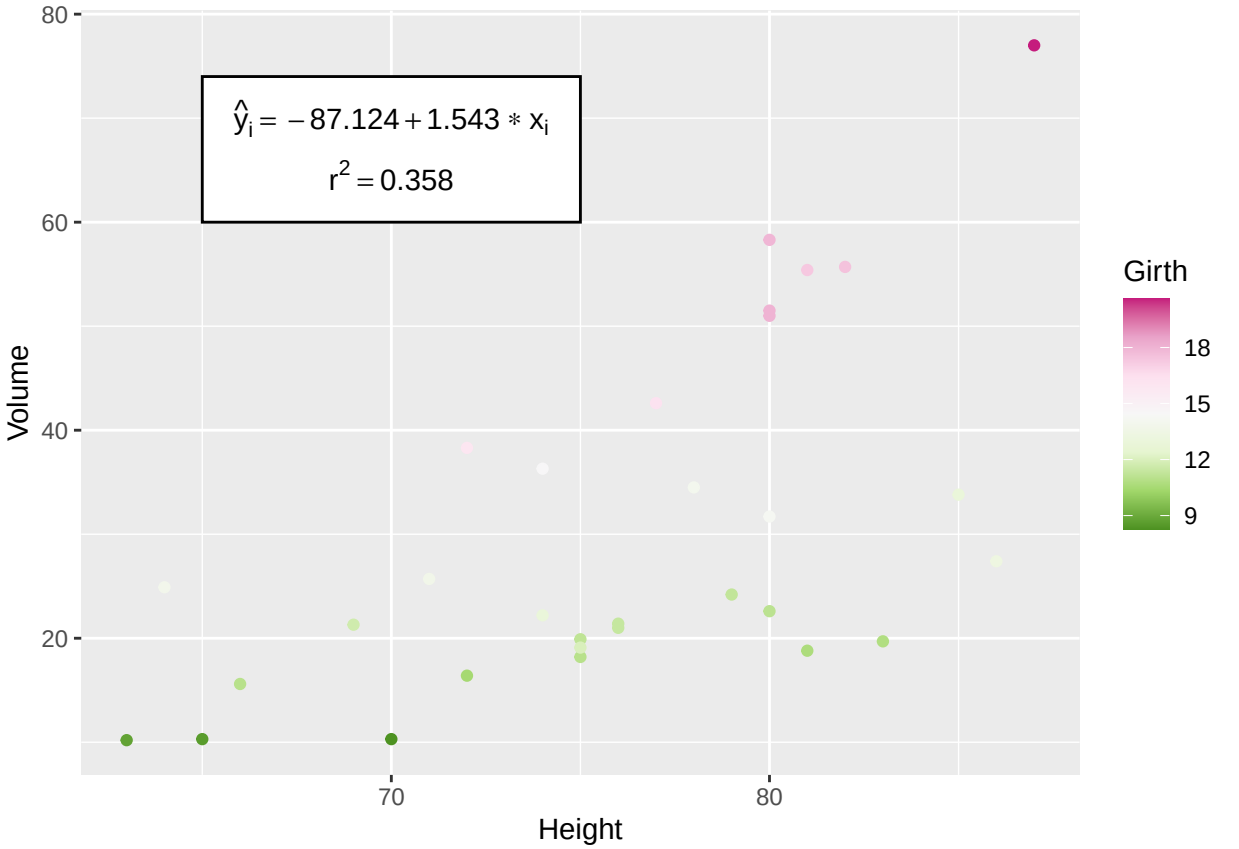
```
summary(model.trees)$r.squared
```

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

```
ggplot(trees,
       aes(x=Height, y = Volume)) +
  geom_point(aes(color=Girth)) +
  scale_color_distiller(palette = 'PiYG') +
  annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74, fill='white', color='black') +
  annotate('text', x = 70, y = 70, label=latex2exp::TeX('$\\hat{y}_i = -87.124 + 1.543* x_i$')) +
  annotate('text', x = 70, y = 64.5, label= latex2exp::TeX('$r^2 = 0.358$'))
```

```
## 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 <- trees %>%
  dplyr::select(-matches('fit'), -matches('lwr'), -matches('upr')) %>%
  cbind( predict(model.trees, newdata=., interval='confidence') )

head(trees)
```

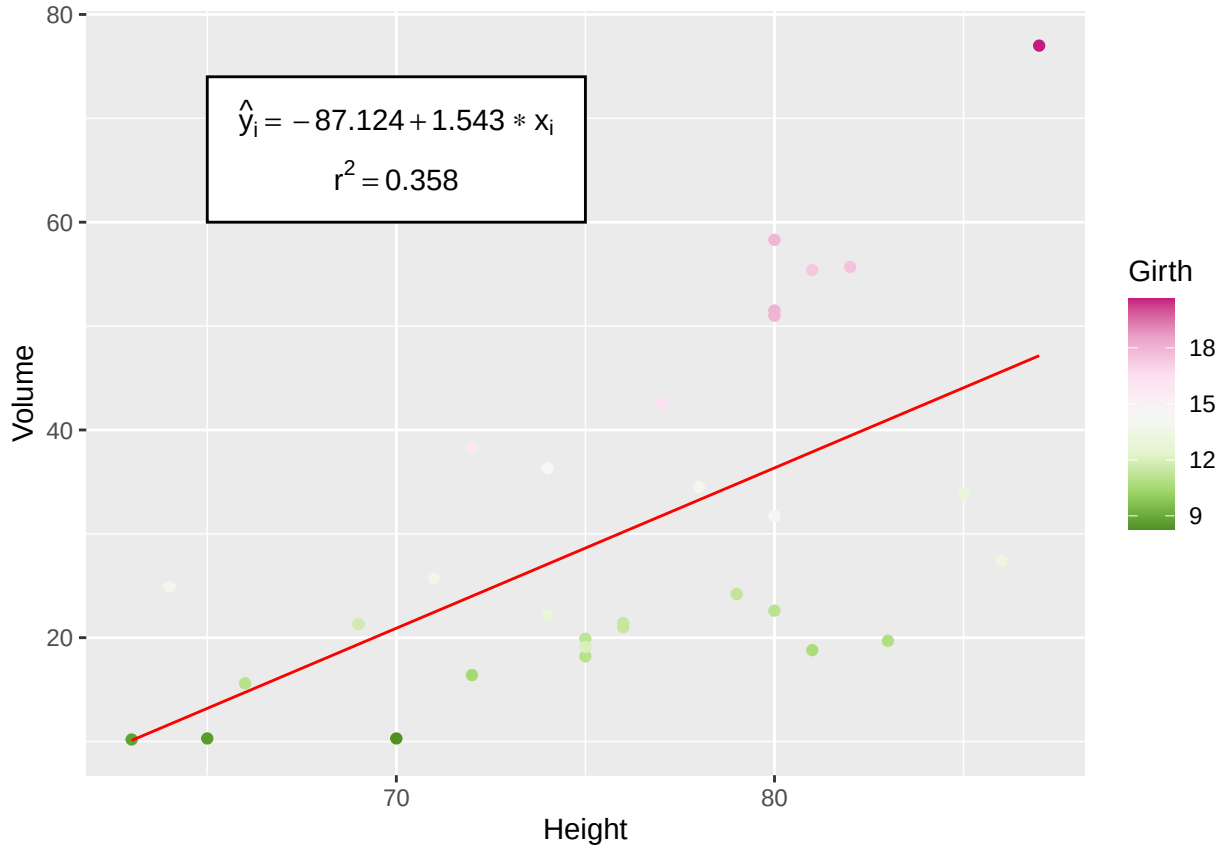
```
##   Girth Height Volume      fit       lwr      upr
## 1   8.3     70   10.3 20.91087 14.098550 27.72319
## 2   8.6     65   10.3 13.19412  3.254288 23.13395
## 3   8.8     63   10.2 10.10742 -1.223363 21.43821
## 4  10.5     72   16.4 23.99757 18.159758 29.83538
## 5  10.7     81   18.8 37.88772 31.592680 44.18275
## 6  10.8     83   19.7 40.97442 33.597379 48.35145
```

```
ggplot(trees,
       aes(x=Height, y = Volume)) +
  geom_point(aes(color=Girth)) +
  geom_line(aes(y = fit), color='red') +
  scale_color_distiller(palette = 'PiYG') +
  annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74, fill='white', color='black') +
  annotate('text', x = 70, y = 70, label=latex2exp::TeX('$\\hat{y}_i = -87.124 + 1.543* x_i$')) +
  annotate('text', x = 70, y = 64.5, label= latex2exp::TeX('$r^2 = 0.358$'))
```

```
## 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'
```

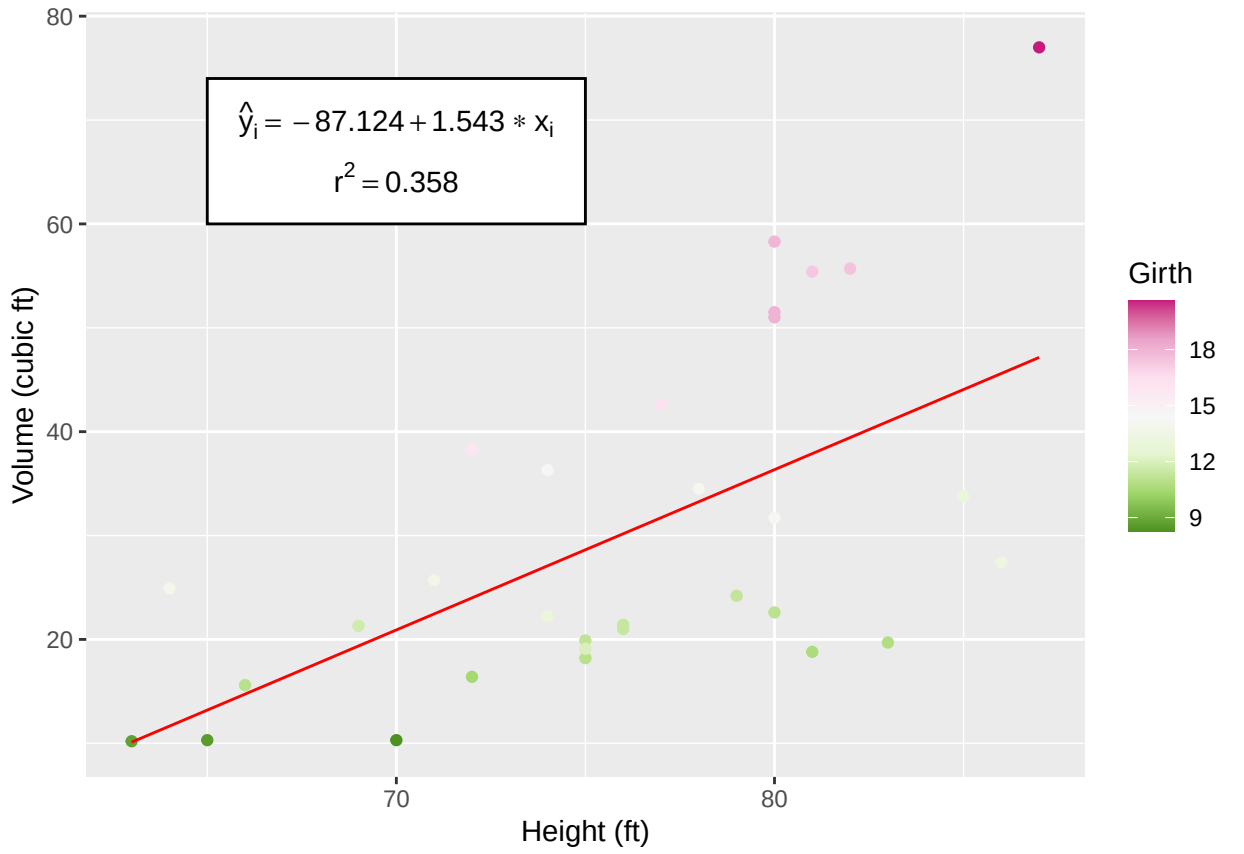


j) Properly label the axes of the graph.

```
ggplot(trees,
       aes(x=Height, y = Volume)) +
  geom_point(aes(color=Girth)) +
  geom_line(aes(y = fit), color='red') +
  scale_color_distiller(palette = 'PiYG') +
  annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74, fill='white', color='black') +
  annotate('text', x = 70, y = 70, label=latex2exp::TeX('$\\hat{y}_i = -87.124 + 1.543* x_i$')) +
  annotate('text', x = 70, y = 64.5, label= latex2exp::TeX('$r^2 = 0.358$')) +
  labs(x = "Height (ft)", y = "Volume (cubic ft)")
```

```
## 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'
```

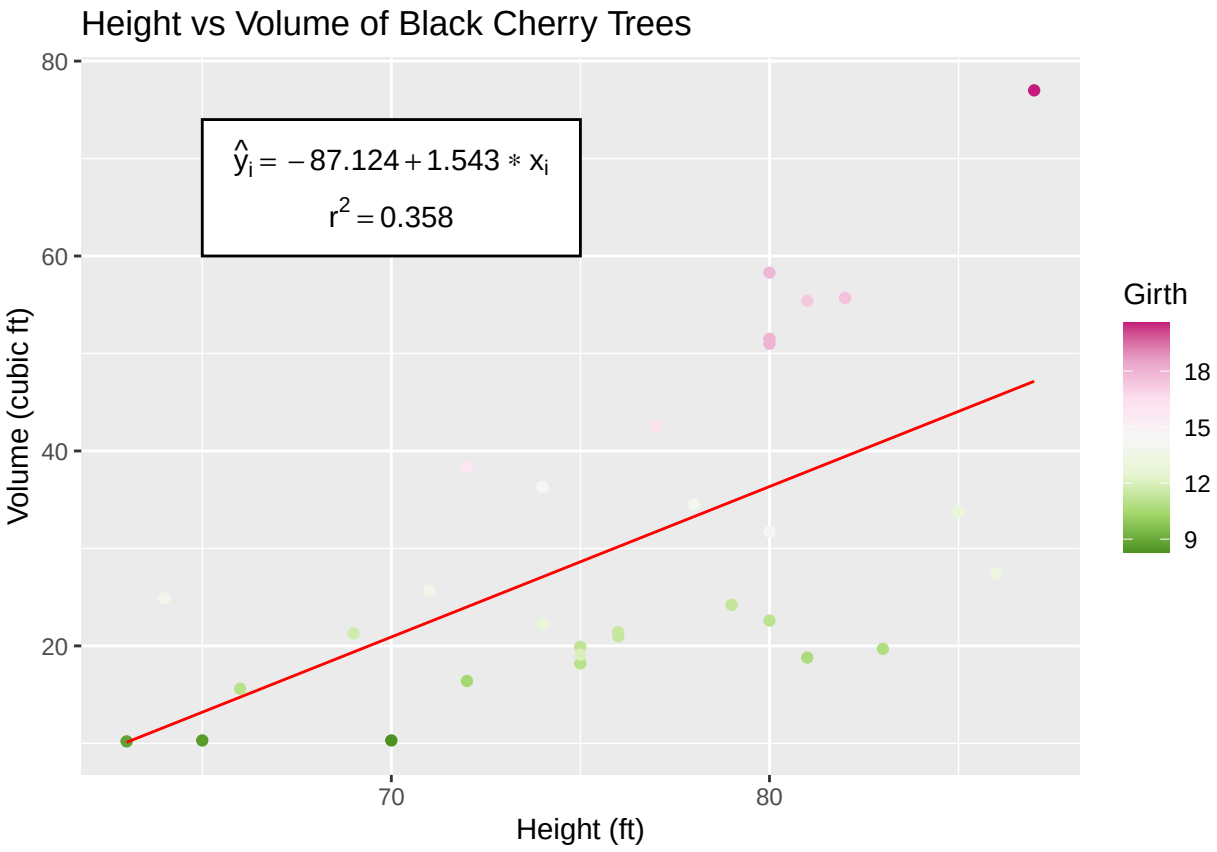


k) Add a descriptive title to the graph.

```
ggplot(trees,
       aes(x=Height, y = Volume)) +
  geom_point(aes(color=Girth)) +
  geom_line(aes(y = fit), color='red') +
  scale_color_distiller(palette = 'PiYG') +
  annotate('rect', xmin=65, xmax=75, ymin=60, ymax=74, fill='white', color='black') +
  annotate('text', x = 70, y = 70, label=latex2exp::TeX('$\\hat{y}_i = -87.124 + 1.543* x_i$')) +
  annotate('text', x = 70, y = 64.5, label= latex2exp::TeX('$r^2 = 0.358$')) +
  labs(x = "Height (ft)", y = "Volume (cubic ft)") +
  labs(title="Height vs Volume of Black Cherry Trees")
```

```
## 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'
```


Height vs Volume of Black Cherry Trees
$\hat{y}_i = -87.124 + 1.543 * x_i$
$r^2 = 0.358$
Volume (cubic ft)
Height (ft)
80
60
40
20
70
80
Girth
18
15
12
9