

assignment15

Noah Plant

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 $y = \text{Volume}$ as a function of $x = \text{Height}$.

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggplot2)
library(broom)

myData<-datasets::trees

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

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

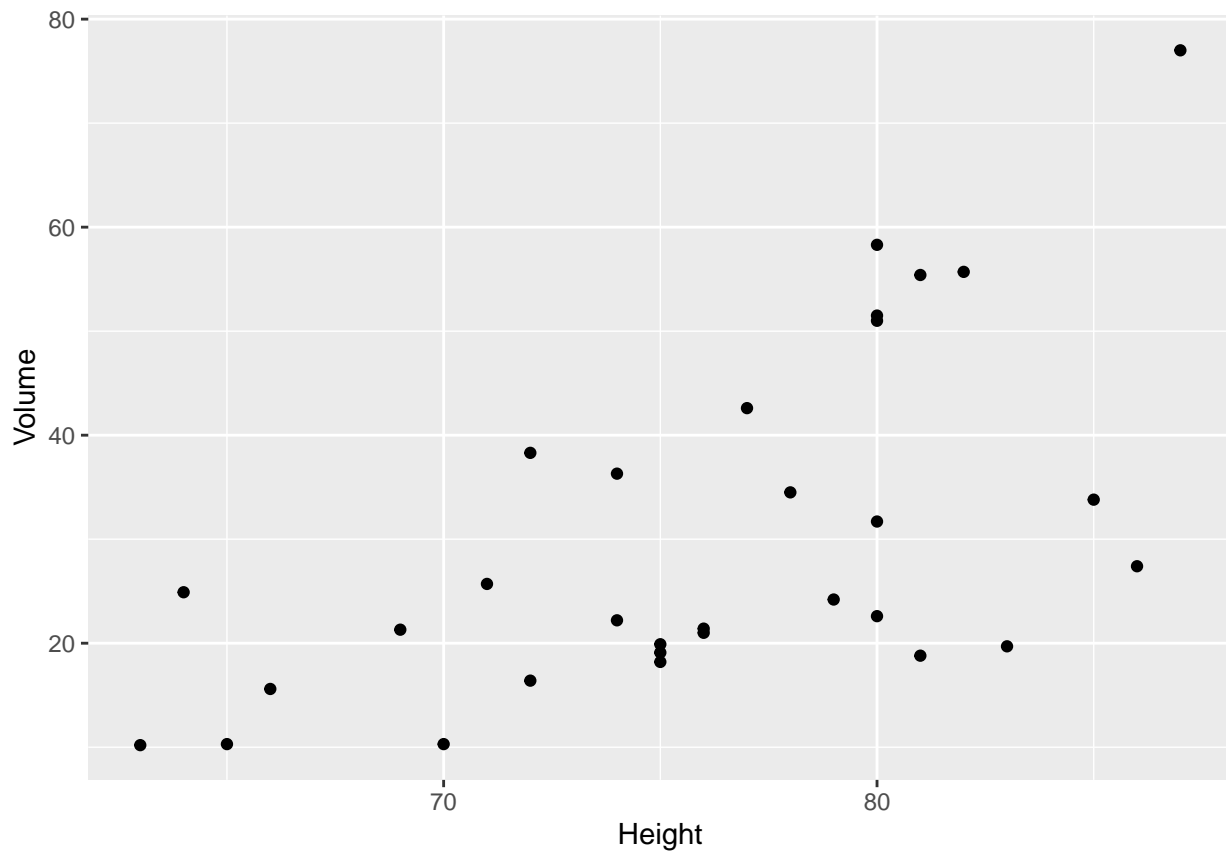
```
summary(myData.model)

##
## Call:
## lm(formula = Volume ~ Height, data = myData)
##
## 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
```

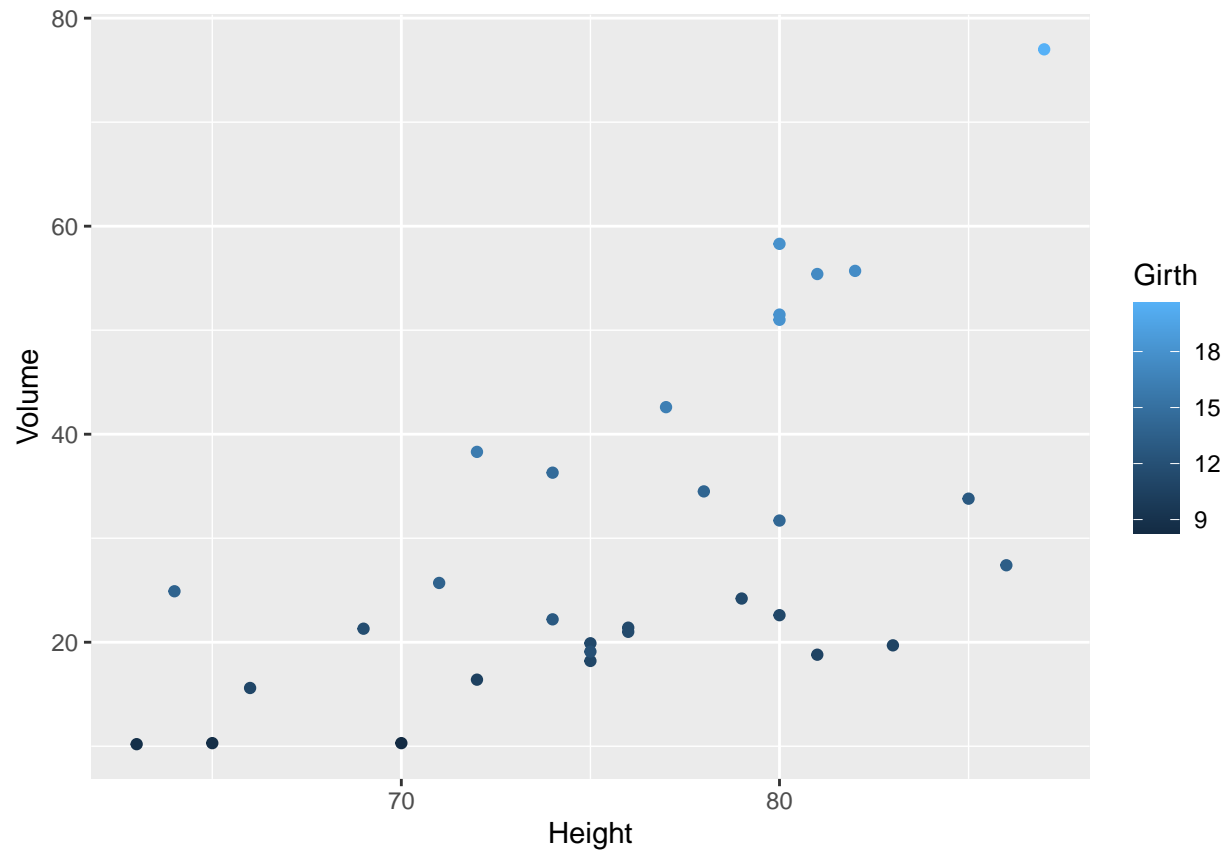
c) Using ggplot2, create a scatter plot of Volume vs Height.

```
plot.c<-myData%>%ggplot(aes(y=Volume,x=Height))+geom_point()
plot.c
```



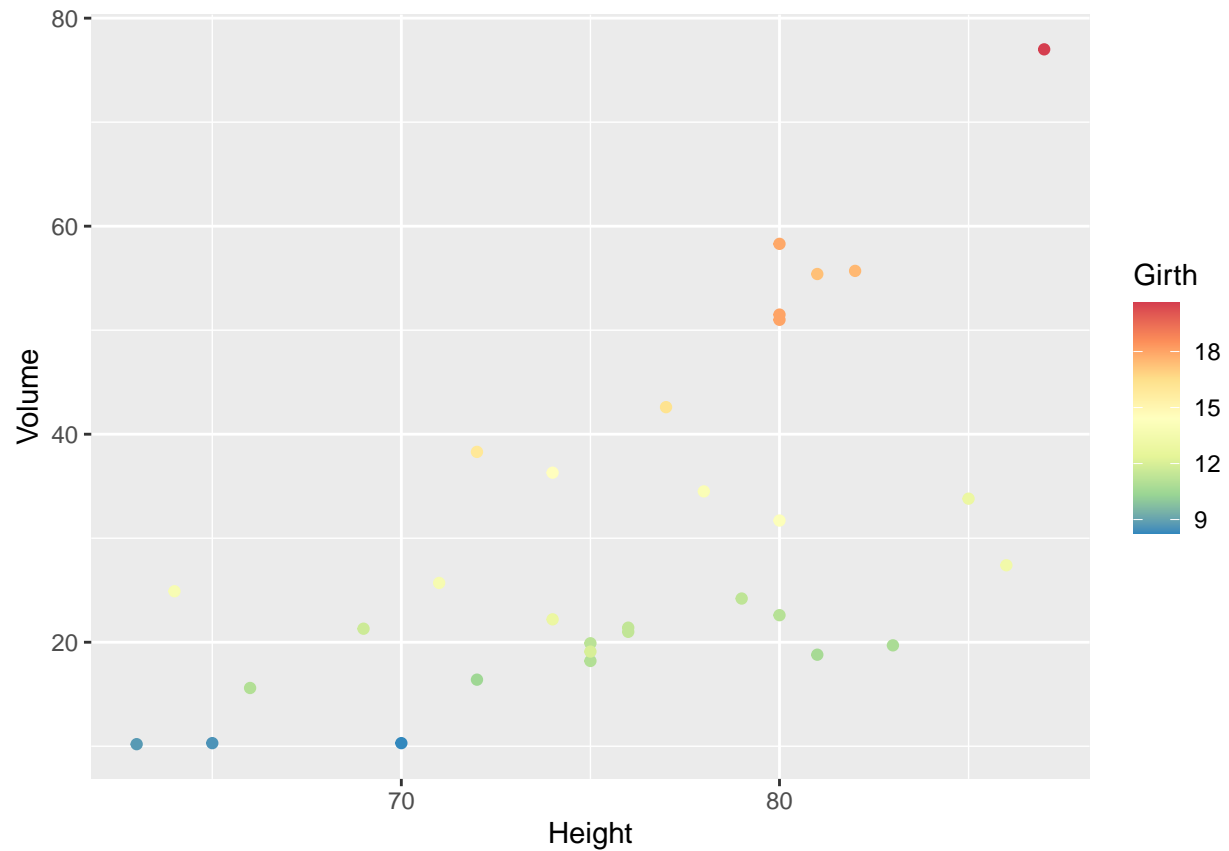
d) Color the scatter using the Girth variable.

```
plot.d<-myData%>%ggplot(aes(y=Volume,x=Height))+geom_point(aes(color=Girth))
plot.d
```



e) Modify the color scheme using a RColorBrewer palette.

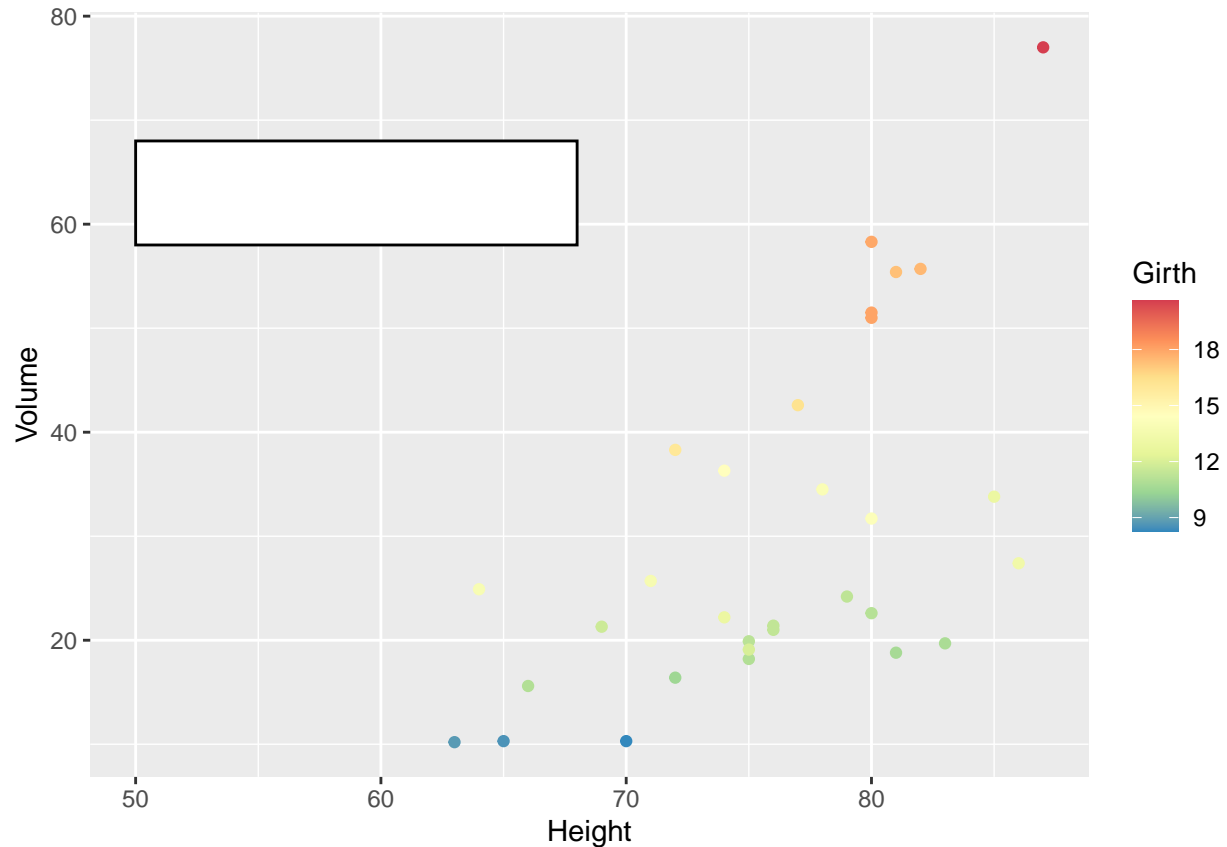
```
plot.e<-plot.d+scale_color_distiller(palette='Spectral')  
plot.e
```



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

```
plot.f<-plot.e+annotate('rect', xmin=50, xmax=68, ymin=58, ymax=68,
  fill='white', color='black')
```

```
plot.f
```



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.

```
library(latex2exp)
```

```
## Warning: package 'latex2exp' was built under R version 4.4.2
```

```
cleanModel<-broom::tidy(myData.model)
```

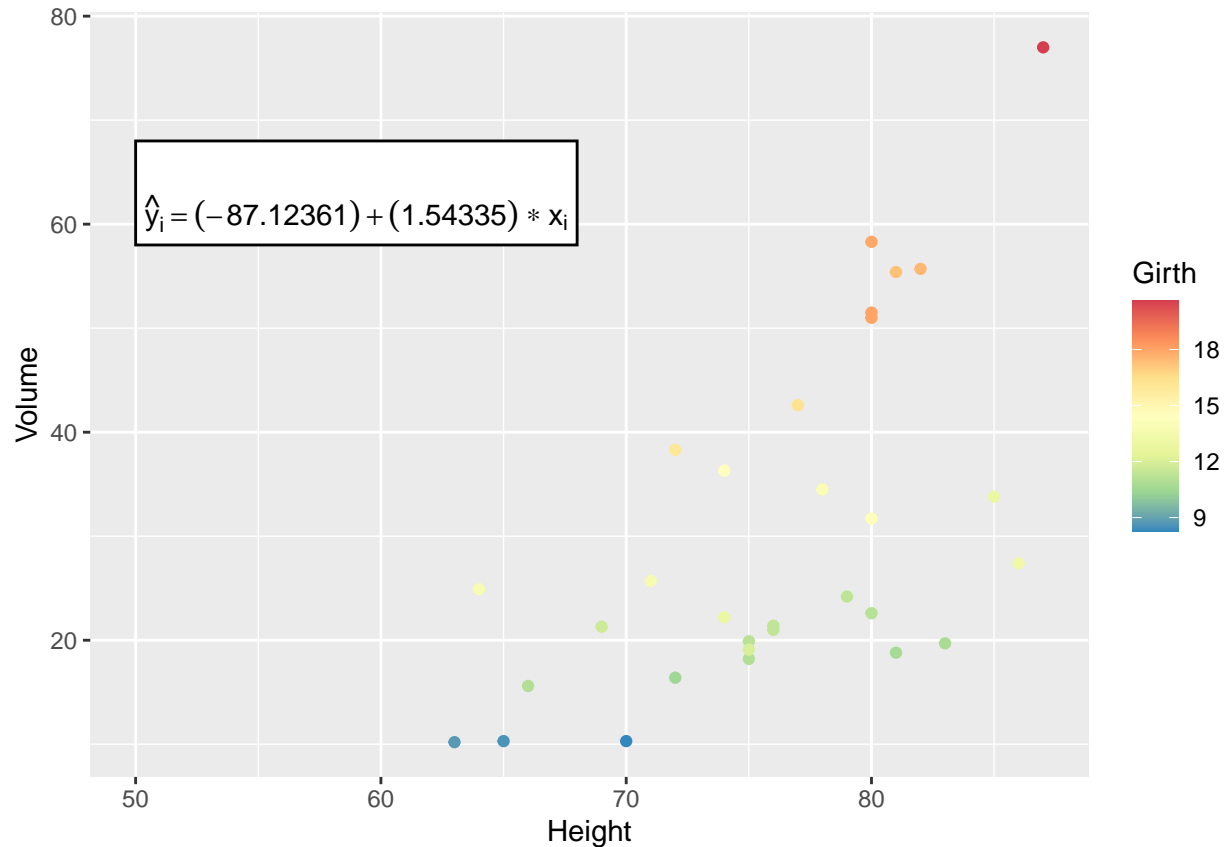
```
intercept<-cleanModel[1,2]
```

```
slope<-cleanModel[2,2]
```

```
plot.g<-plot.f+annotate("text",x=59,y=61,label=latex2exp::TeX('$\\hat{y}_i = (-87.12361 ) + (1.54335
```

```
plot.g
```

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

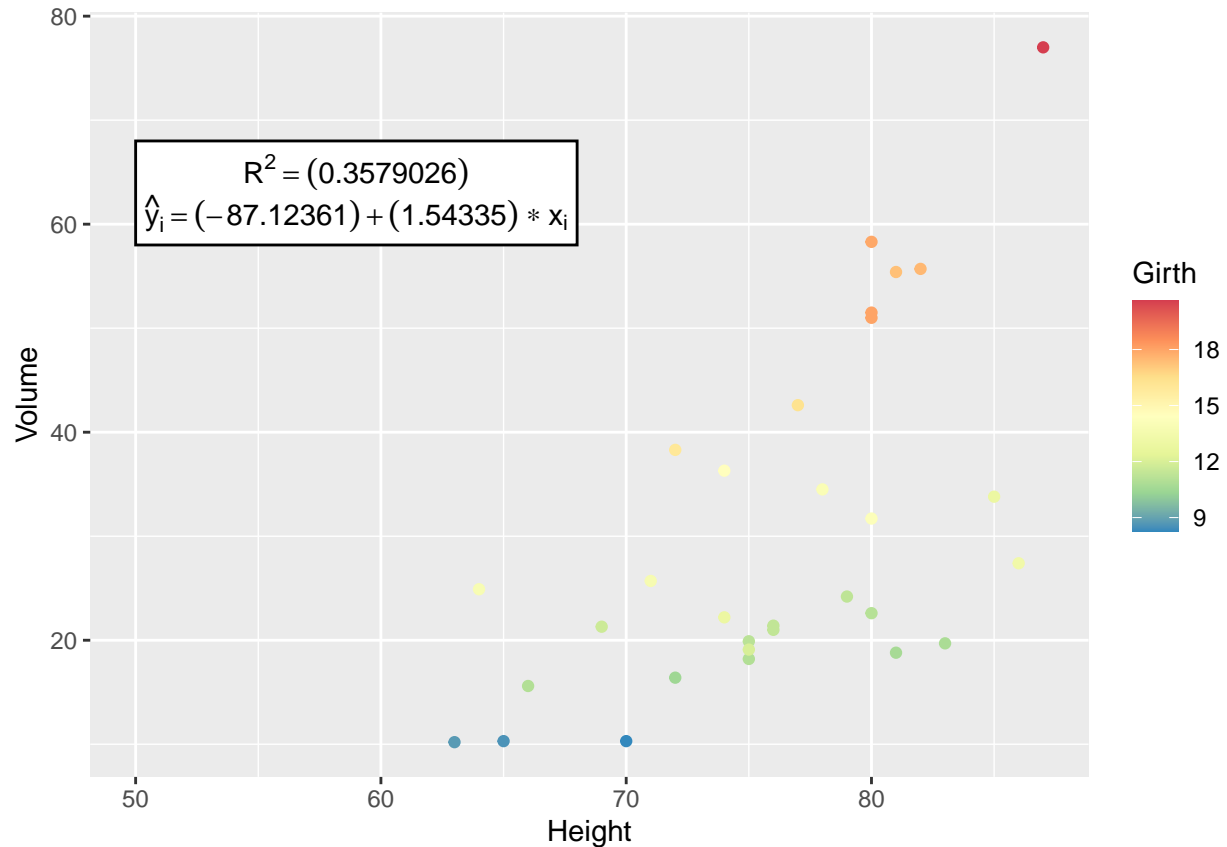
```
rsquared<-broom::glance(myData.model)

rsquared<-rsquared[1]

plot.h<-plot.g+annotate('text',x=59,y=65,label=latex2exp::TeX('$R^2 = (0.3579026 )$'))

plot.h
```

```
## 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.

```
line<-broom::augment(myData.model, trees)
```

```
line
```

```
## # A tibble: 31 x 9
##   Girth Height Volume .fitted .resid .hat .sigma .cooksd .std.resid
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  8.3    70   10.3  20.9 -10.6  0.0618  13.5  0.0220  -0.818
## 2  8.6    65   10.3  13.2  -2.89  0.132   13.6  0.00407  -0.232
## 3  8.8    63   10.2  10.1  0.0926  0.171   13.6  0.00000594  0.00759
## 4 10.5    72   16.4  24.0  -7.60  0.0454  13.6  0.00801  -0.580
## 5 10.7    81   18.8  37.9 -19.1  0.0528  13.1  0.0597  -1.46
## 6 10.8    83   19.7  41.0 -21.3  0.0725  13.0  0.106  -1.65
## 7 11      66   15.6  14.7  0.863  0.114   13.6  0.000302  0.0684
## 8 11      75   18.2  28.6 -10.4  0.0331  13.5  0.0107  -0.792
## 9 11.1    80   22.6  36.3 -13.7  0.0454  13.4  0.0262  -1.05
## 10 11.2    75   19.9  28.6  -8.73  0.0331  13.5  0.00751  -0.663
## # i 21 more rows
```

```
plot.i<-line%>%ggplot(aes(y=Volume,x=Height))+
  geom_point(aes(color=Girth))+
  geom_line(aes(y=.fitted))+
  scale_color_distiller(palette='Spectral')+
  annotate('rect', xmin=50, xmax=68, ymin=58, ymax=68,
```

```

    fill='white', color='black')+
  annotate("text",x=59,y=61,
    label=latex2exp::TeX('$\\hat{y}_i = (-87.12361 ) + (1.54335 ) * x_i$'))+
  annotate("text",x=59,y=65,label=latex2exp::TeX('$R^2 = (0.3579026 )$'))

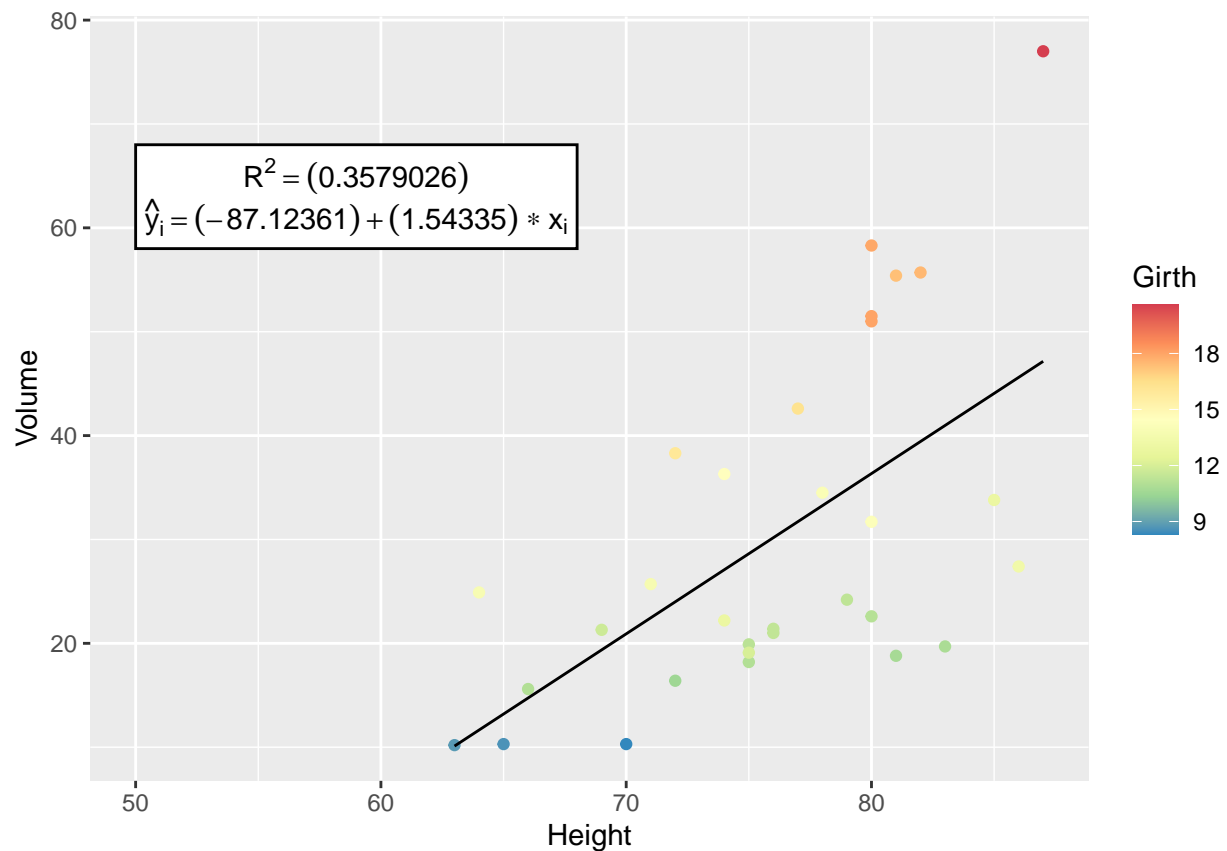
```

plot.i

```

## 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.

```

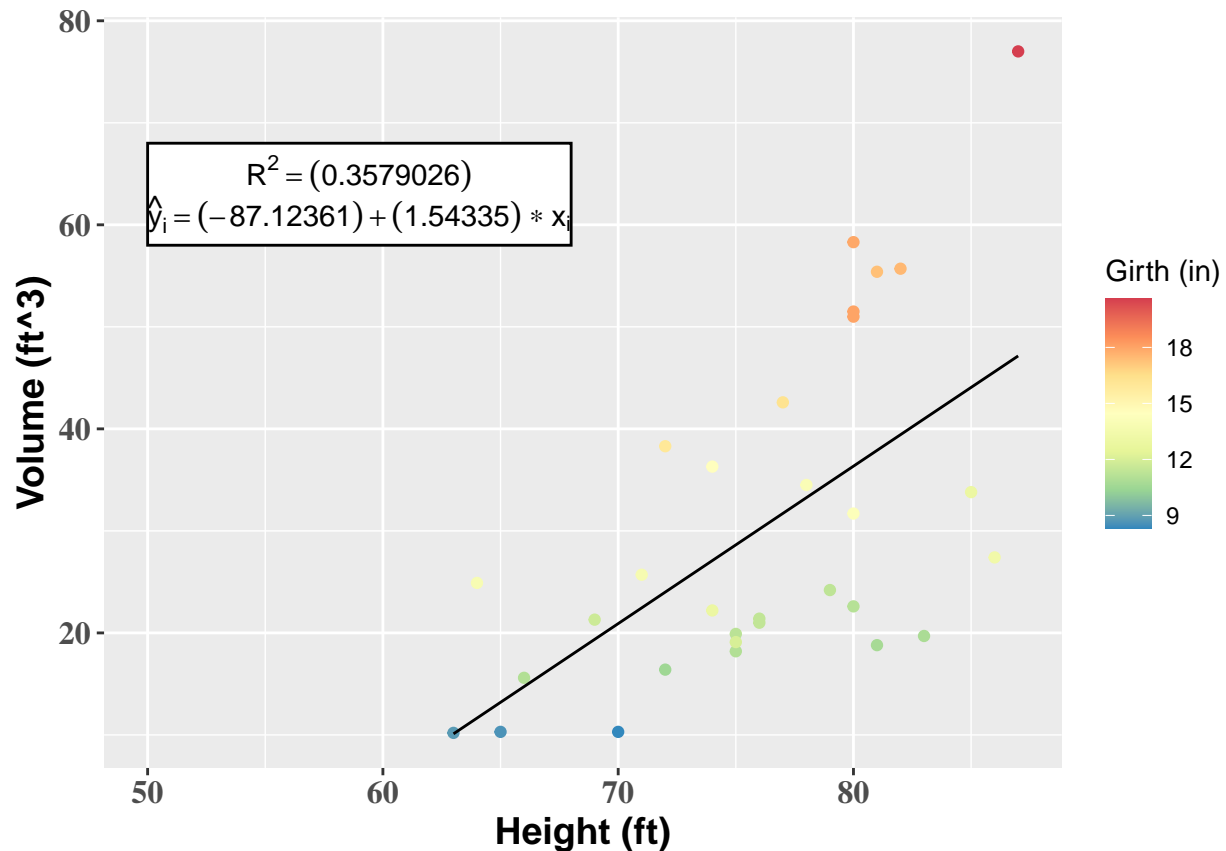
plot.j<-plot.i+ labs(x= "Height (ft)",y="Volume (ft^3)",color="Girth (in)")+theme(
  # Change x-axis font type
  axis.text.x = element_text(size = 12, face = "bold", family = "serif"),
  # Change y-axis font type
  axis.text.y = element_text(size = 12, face = "bold", family = "serif"),
  # Change x-axis title font type
  axis.title.x = element_text(size = 14, face = "bold"),
  # Change y-axis title font type

```



```
axis.title.y = element_text(size = 14, face = "bold")
)
plot.j
```

```
## 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.

```
plot.k<-plot.j+labs(title="Black Cherry Tree Dimensions")
plot.k
```

```
## 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'
```

