

Homework 3 - Coding

John Kaspers

May 27, 2020

Warning

Warning: You must have the “PPforest” package and “effects” package installed for this assignment. You can install these packages by running the following code in base R (not RStudio or RMarkdown).

```
install.packages("PPforest")
install.packages("effects")
```

Data

The original data set contains measurements of 159 fish from 7 species. Altogether there are 7 variables. All of the fish are caught from the same lake (Laengelmaves) near Tampere in Finland.

- **Type:** has 7 fish classes (Bream, Parkki, Perch, Pike, Roach, Smelt, and Whitewish).
- **weight:** Weight of the fish (in grams)
- **length1:** Length from the nose to the beginning of the tail (in cm)
- **length2:** Length from the nose to the notch of the tail (in cm)
- **length3:** Length from the nose to the end of the tail (in cm)
- **height:** Height of the fish (in cm)
- **width:** Width of the fish (in cm)

Two data sets have been created from this original data and will be used for your analysis.

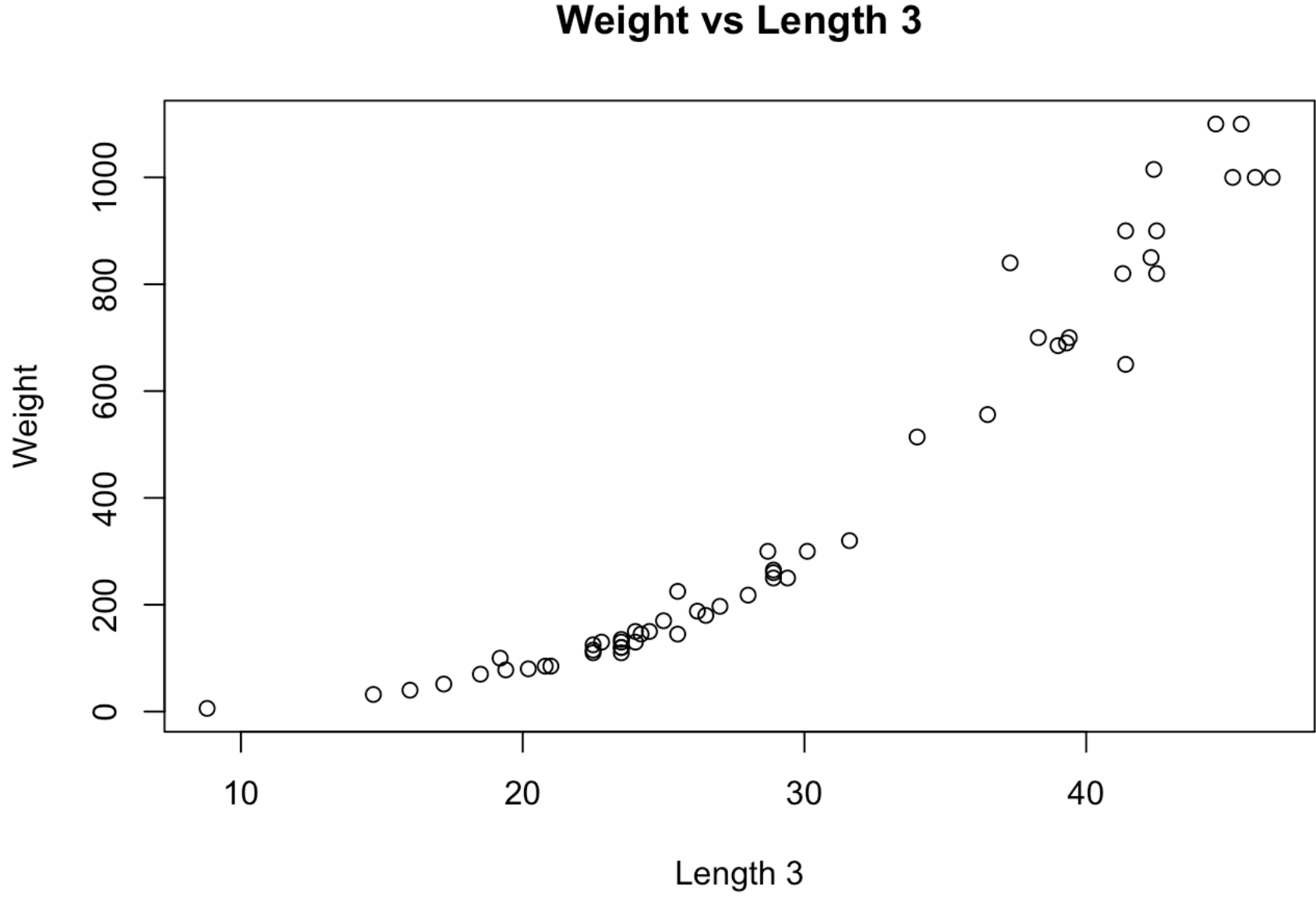
- The **perch** data set has been filtered to include only data on perch (56 observations)
- The **fish** data set has been filtered to include only data on perch, bream, and pike (108 observations)

perch

In this section, we will utilize the **perch** data set that was created above. Be sure to run the 'setup' and 'data' code blocks before completing this section. Here, we look to predict weight based off the overall length of a fish for the perch species.

1. Create a scatterplot of `weight` vs `length3`. Give appropriate labels for your axes.

```
# Write your code here
plot(perch$length3, perch$weight, xlab = "Length 3", ylab = "Weight", main = "Weight vs Length 3")
```



2. Does the relationship appear to be linear or non-linear (curved)? If it appears linear, estimate (eyeball) the slope. If it appears non-linear, what should we include in our model to better predict the response?

This relationship appears to be quadratic. Perhaps if we were to include `length1` and `length2` in our model, or maybe squaring `length3` and including both a linear term and a square term we could better predict weight.

3. Using the information from Parts 1 and 2, create a model to *best* predict `weight` using `length3` and run it through the `summary()` function.

```
lm.quadratic <- lm(weight~length3+I(length3^2), data = perch)
summary(lm.quadratic)
```

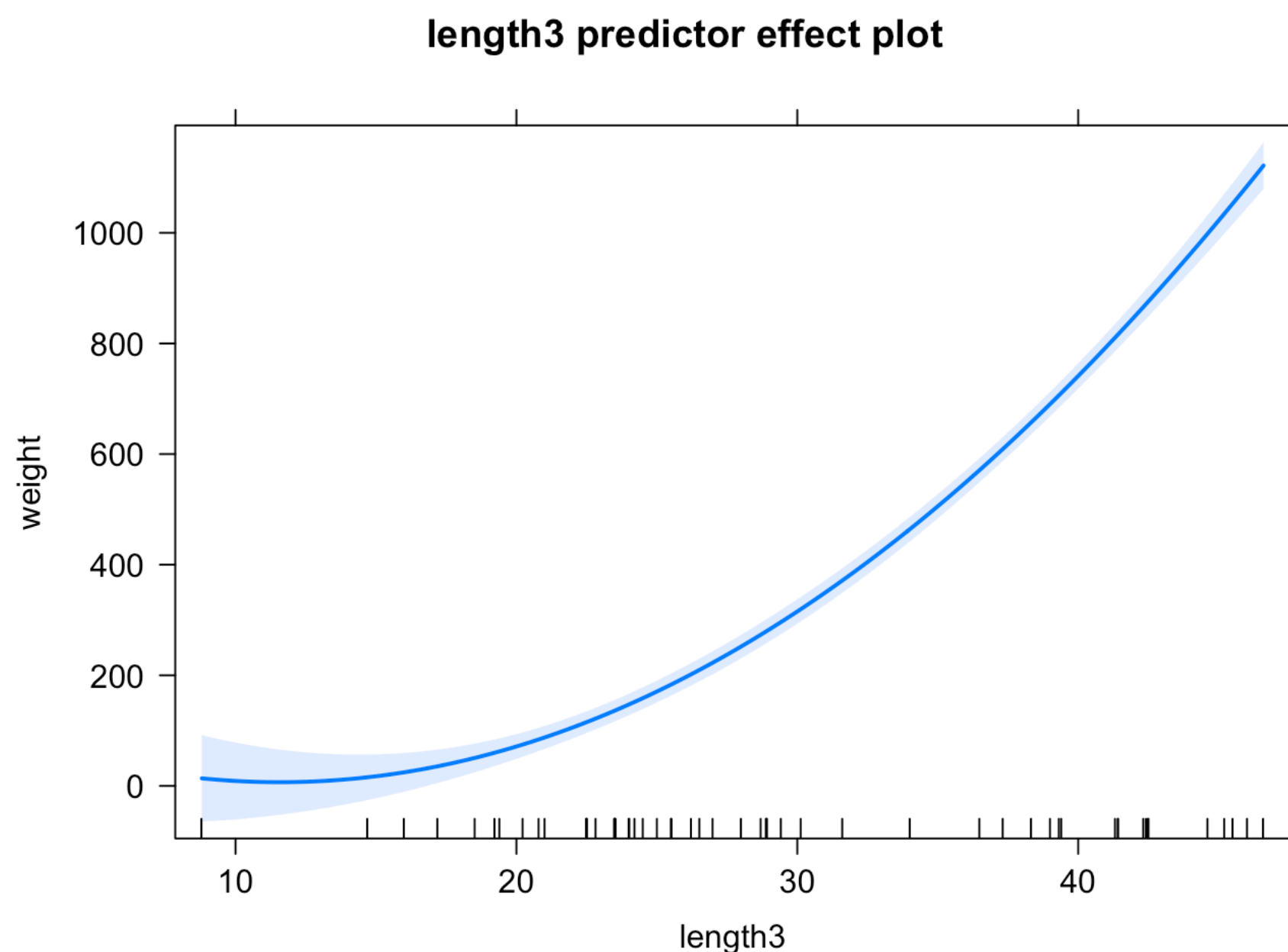
```
##
## Call:
## lm(formula = weight ~ length3 + I(length3^2), data = perch)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -165.295  -18.970   -5.747   15.652   231.691
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  128.34533    78.77870     1.629   0.10920
## length3     -21.02388     5.41770    -3.881   0.00029 ***
## I(length3^2)   0.90862     0.08689   10.458 1.72e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 56.99 on 53 degrees of freedom
## Multiple R-squared:  0.9741, Adjusted R-squared:  0.9731
## F-statistic: 996.6 on 2 and 53 DF,  p-value: < 2.2e-16
```

4. Comment on the statistical significance (using a 5% significance level) of the predictor(s) in your model.

Because the p-value for `length3`, given by 0.00029, and the p-value for `I(length3^2)`, given by 1.72×10^{-14} , are both well below 0.05 we can conclude that the model is statistically significant at the 5% significance level.

5. Visualize the model you created. Pass your model into the following code. Note: Do not worry about changing the default labels for this plot.

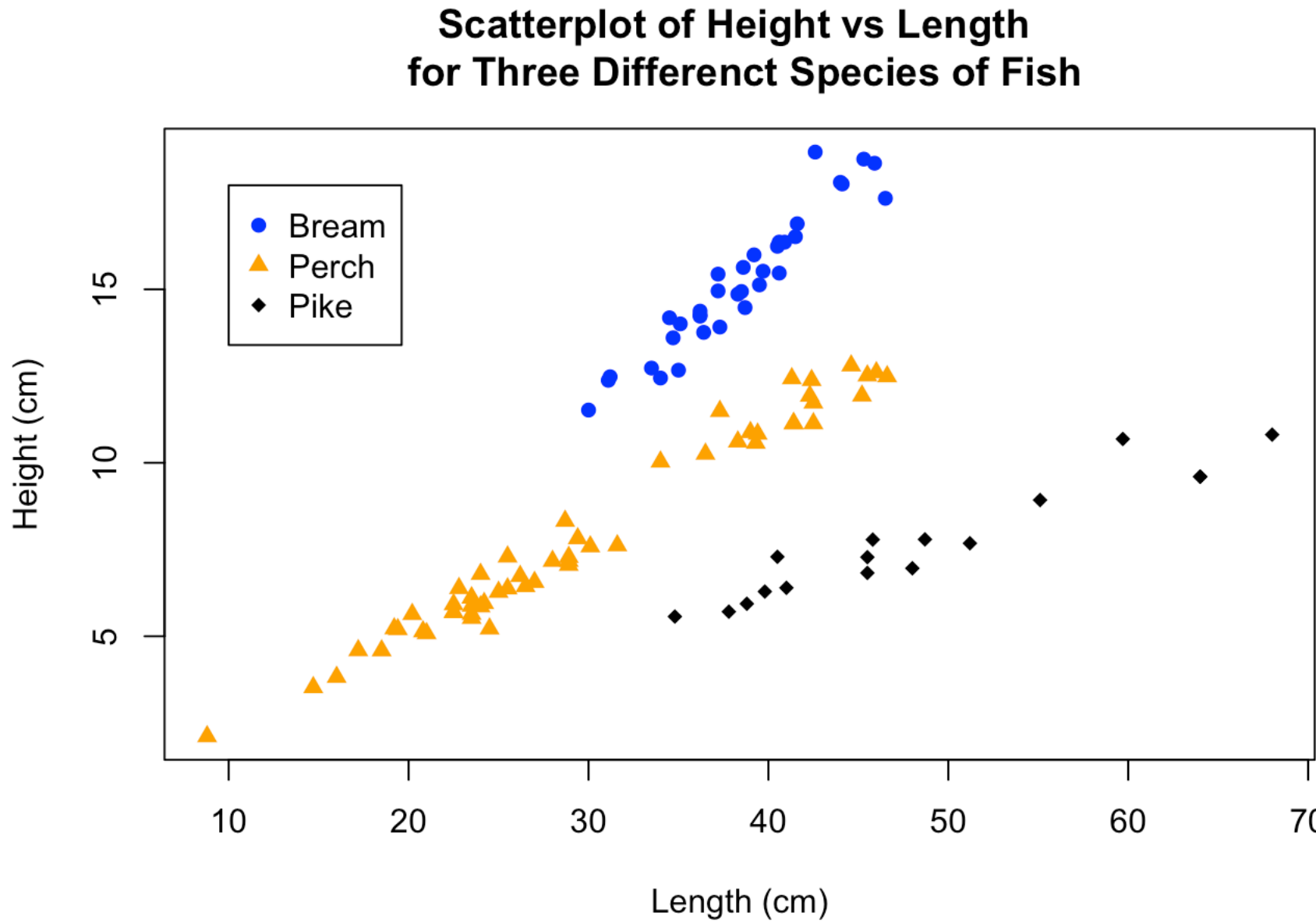
```
# Uncomment the following line of code and change 'lm.model' to the name of your model
plot(predictorEffects(lm.quadratic))
```



fish

In this section, we will utilize the **fish** data set that was created above. Be sure to run the 'setup' and 'data' code blocks before completing this section. Here, we look to predict height based off the overall length of a fish for three different species.

Run the following code block to visualize the relationship between `height` and `length3` for the three different species.



1. Briefly explain why this data would benefit from an interaction term between length and species.

This data would benefit from an interaction term between length and species because the data show pretty similar slopes for the three species. This, coupled with the fact that the data are grouped by species, makes introducing an interaction term a good idea.

2. Create the model to predict `height` using `length3` and `Type`. Include an interaction term between the two predictors. Run your model through the `summary()` function.

```
lm.fish.interaction <- lm(height~length3 + Type + length3 * Type, data = fish)
summary(lm.fish.interaction)
```

```
##
## Call:
## lm(formula = height ~ length3 + Type + length3 * Type, data = fish)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22750  -0.31983   0.01588   0.30107   1.86206
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -2.08664     0.84661   -2.465   0.0154 *
## length3         0.45027     0.02195   20.515 < 2e-16 ***
## TypePerch       1.14296     0.87828   1.301   0.1961
## TypePike        2.20032     1.06757   2.061   0.0418 *
## length3:TypePerch -0.15250     0.02320  -6.572 2.14e-09 ***
## length3:TypePike  -0.29427     0.02555 -11.516 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5321 on 102 degrees of freedom
## Multiple R-squared:  0.9849, Adjusted R-squared:  0.9841
## F-statistic: 1330 on 5 and 102 DF,  p-value: < 2.2e-16
```

3. Comment on the statistical significance (using a 5% significance level) of the two interaction terms in your model.

Both of the interaction terms, `length3:TypePerch`, and `length3:TypePike`, are statistically significant at the 5% level. Additionally `length3` and `TypePike` are statistically significant at the 0.05 level. Also, all of the predictors are statistically significant at the 0.05 level except for `TypePerch`.

4. Calculate the intercept and slope for the relationship between height and length for each of the three species. (Note: You may calculate these by hand, in R, or with a calculator. You may round the estimates from the R output to 2 decimal places when carrying out the calculations.)

- Bream: Intercept = -2.09, Slope = 0.45
- Perch: Intercept = -0.94, Slope = 0.30
- Pike: Intercept = 0.11, Slope = 0.16

5. Visualize the model you created. Pass your model into the following code. Note: Do not worry about changing the default labels for this plot.

```
# Uncomment the following line of code and change 'lm.model' to the name of your model
plot(predictorEffects(lm.fish.interaction, "length3"), multiline = TRUE)
```

