# Fish Market

Karl Mbouombouo

2023-08-21

## Packages

```
library(readr)
library("ggplot2")
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("broom")
library("ggpubr")
library("psych")
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
library('ggiraphExtra')
```

## Data

```
fish.data <- read_csv("~/Desktop/R-Projects/Multiple Linear Regression/Fish Market/Fish.csv")
```

```
## Rows: 159 Columns: 7
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Species
## dbl (6): Weight, Length1, Length2, Length3, Height, Width
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
which(is.na(fish.data))
```

```
## integer(0)
```

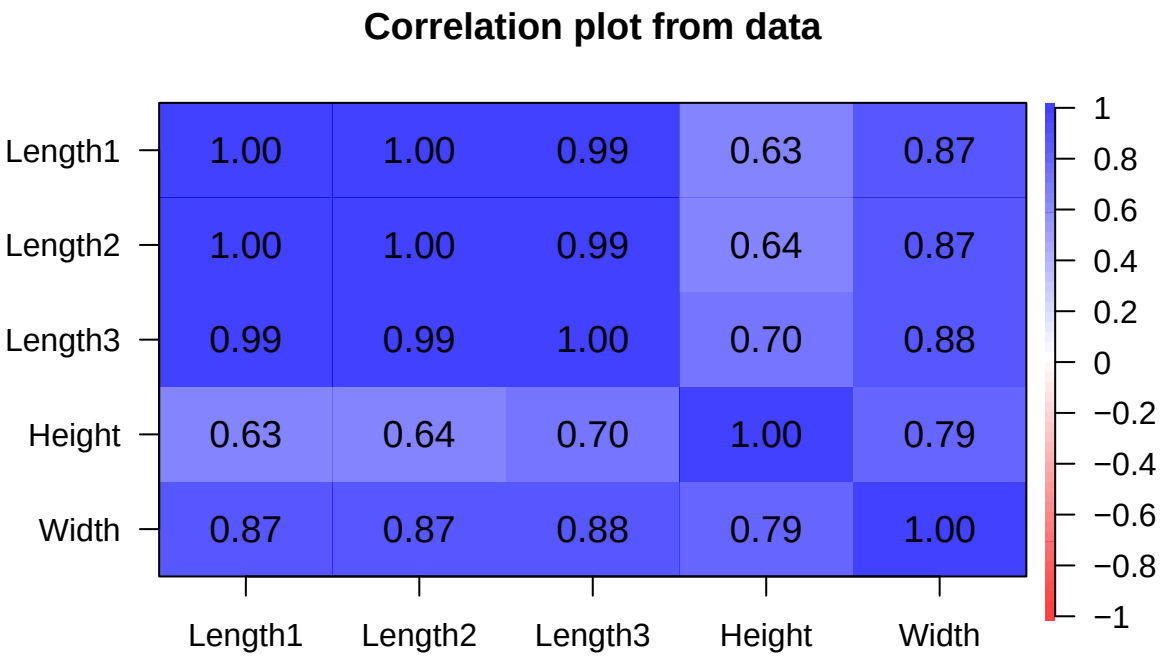
```r
summary(fish.data)
```

```
##    Species              Weight          Length1         Length2
##  Length:159         Min.   :   0.0   Min.   : 7.50   Min.   : 8.40
##  Class :character   1st Qu.: 120.0   1st Qu.:19.05   1st Qu.:21.00
##  Mode  :character   Median : 273.0   Median :25.20   Median :27.30
##                     Mean   : 398.3   Mean   :26.25   Mean   :28.42
##                     3rd Qu.: 650.0   3rd Qu.:32.70   3rd Qu.:35.50
##                     Max.   :1650.0   Max.   :59.00   Max.   :63.40
##     Length3          Height           Width
##  Min.   : 8.80   Min.   : 1.728   Min.   :1.048
##  1st Qu.:23.15   1st Qu.: 5.945   1st Qu.:3.386
##  Median :29.40   Median : 7.786   Median :4.248
##  Mean   :31.23   Mean   : 8.971   Mean   :4.417
##  3rd Qu.:39.65   3rd Qu.:12.366   3rd Qu.:5.585
##  Max.   :68.00   Max.   :18.957   Max.   :8.142
```

```r
dim(fish.data)
```

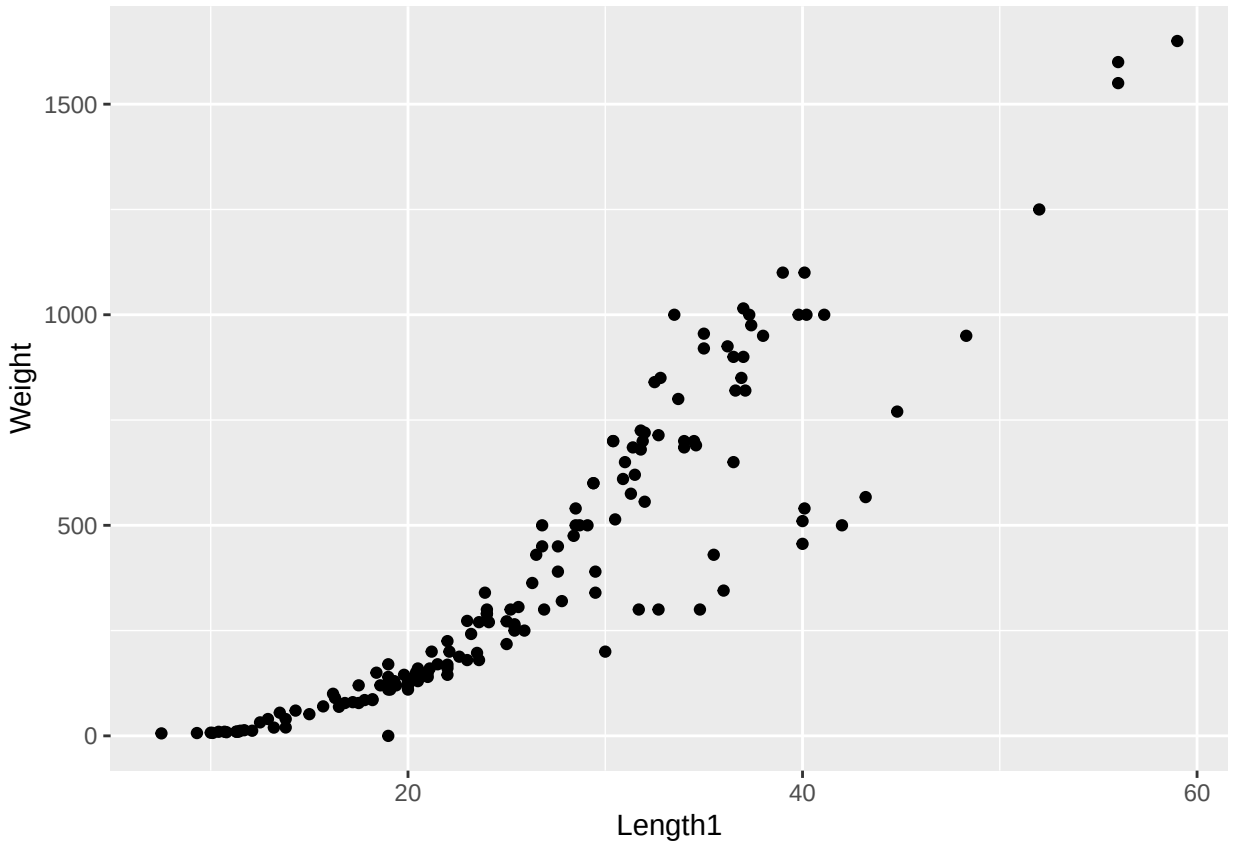
```
## [1] 159   7
```

### Correlation and Plots

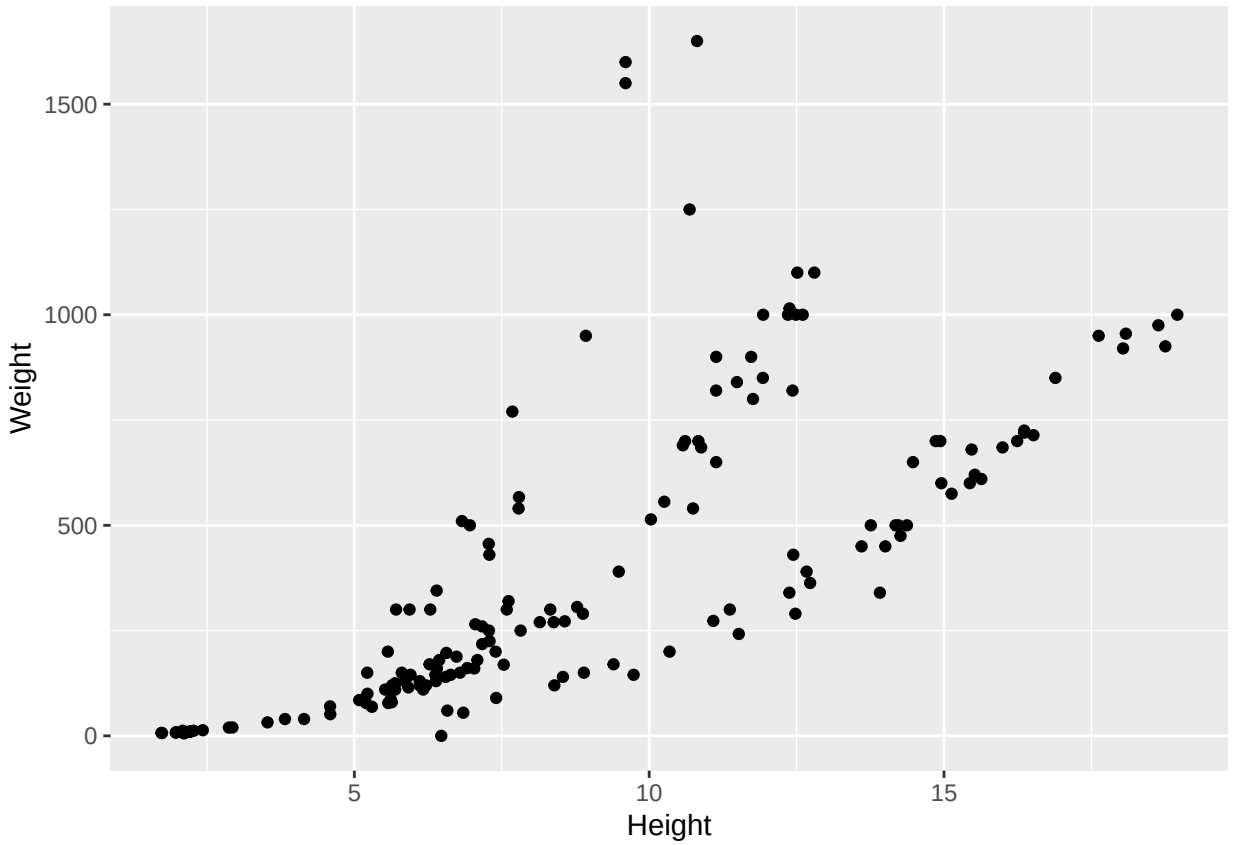



An highly correlation exist between Length1, Length2, and Length3, so we cannot include those parameters together in the model.

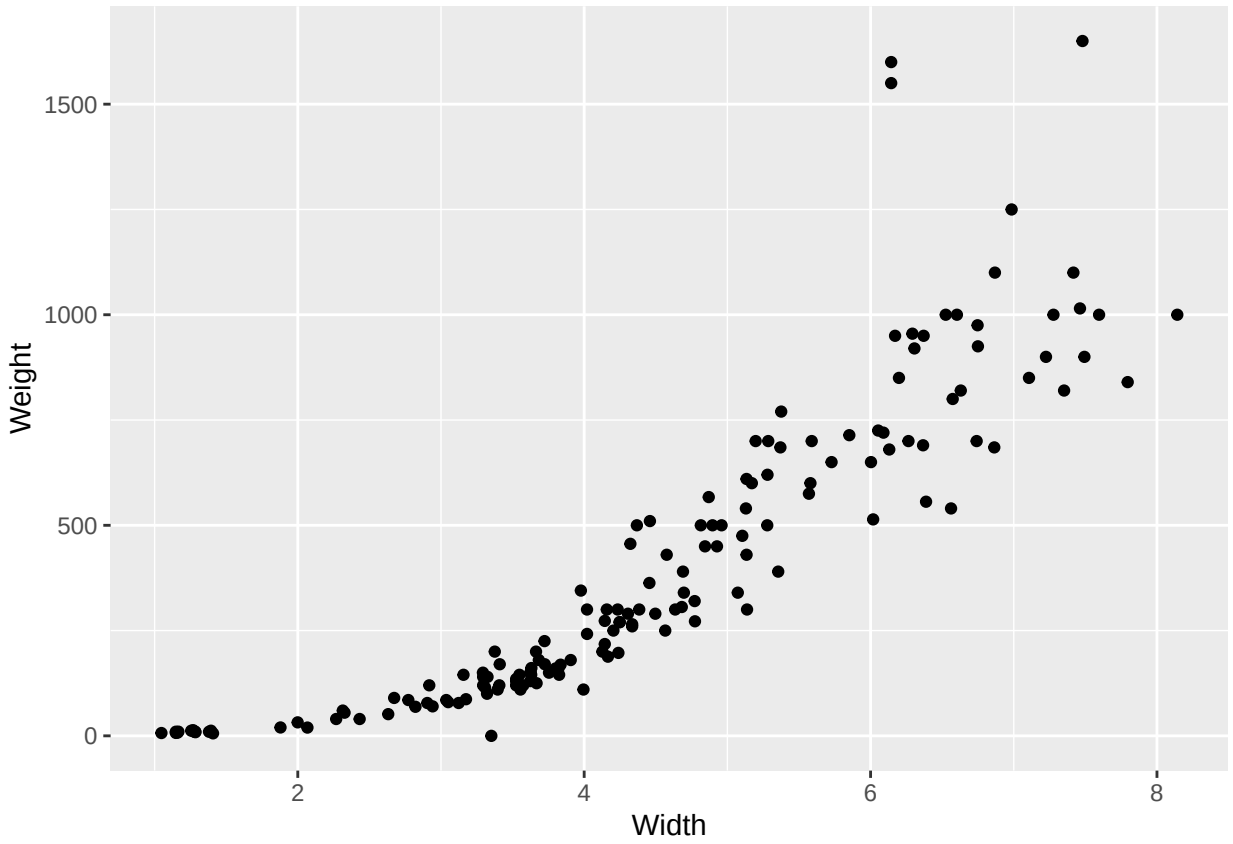
```r
# Relationship between Weight and Length1
ggplot(fish.data, aes(x = Length1, y = Weight)) +
  geom_point()
```



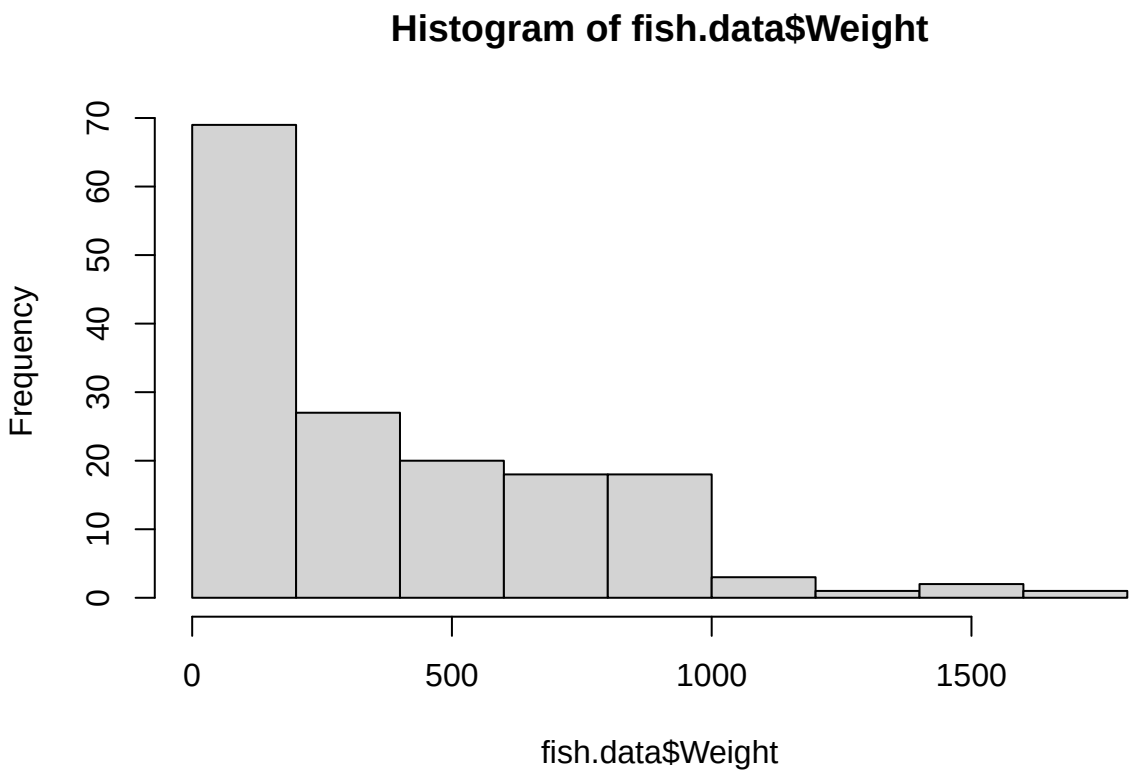

```
# Relationship between Weight and Height
ggplot(fish.data, aes(x = Height, y = Weight)) +
  geom_point()
```




```
# Relationship between Weight and Width
ggplot(fish.data, aes(x = Width, y = Weight)) +
  geom_point()
```




```
hist(fish.data$Weight)
```




The distribution of observations is roughly right-skewed, the mean is greater than the median.

## Modeling

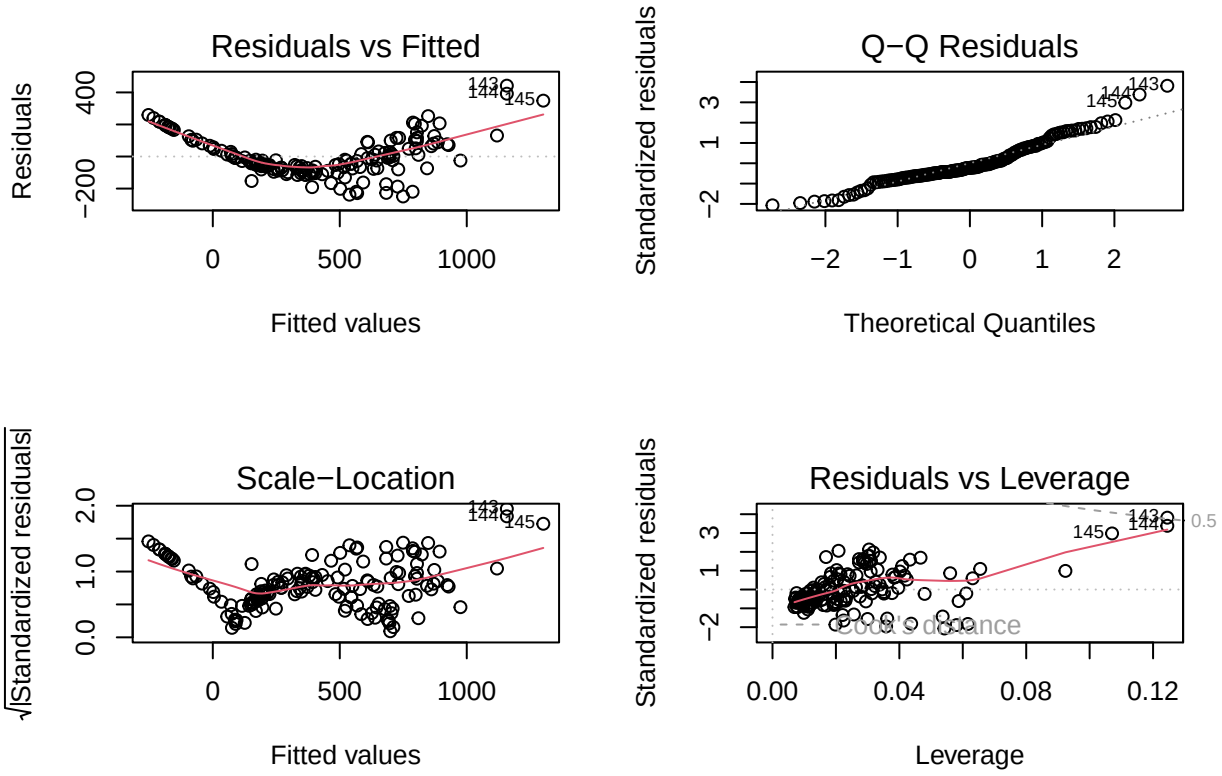
```
fit <- lm(Weight ~ Length1 + Height + Width, data = fish.data)
summary(fit)
```

```
##
## Call:
## lm(formula = Weight ~ Length1 + Height + Width, data = fish.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -249.24  -73.32  -28.17   74.00  442.11
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -514.342     28.549 -18.016  < 2e-16 ***
## Length1       22.683      2.021  11.222  < 2e-16 ***
## Height        13.670      3.854   3.547 0.000516 ***
## Width         44.069     15.347   2.871 0.004659 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 123.9 on 155 degrees of freedom
## Multiple R-squared:  0.8826, Adjusted R-squared:  0.8803
## F-statistic: 388.3 on 3 and 155 DF,  p-value: < 2.2e-16
```

The estimated effect of the length1 on weight is 22.683, while the estimated effect of height is 13.670, and width is 44.069. The p-value here it's 2.2e-16 (or almost zero), which will indicate whether the model fits the data well. We can say that there is a significant positive relationship between Weight and Lenght1, Height and Width, with respectively 22.683 unit, 13.670 unit and 44.069 increase in fish's weight for every unit increase in Lenght1, Height and Width.

## Result visualization

```
par(mfrow=c(2,2))
plot(fit)
```




```
par(mfrow=c(1,1))
```

The residuals are all basically horizontal and centered around zero. This means there are no outliers or biases in the data that would make a linear regression invalid. From the Normal Q-Q plot in the top right, we can see that the real residuals from our model form an almost perfectly one-to-one line with the theoretical residuals from a perfect model. Based on these residuals, we can say that our model meets the assumption of homoscedasticity.

## Conclusion

In our fish surveys, we found significant relationships between the fish weight and the length, height and width ($p < 0$, $p < 0.001$, and $p < 0.005$ respectively). Specifically we found that 44.069 increase ($\pm$ 15.347) in weight (gramme) for every 1cm increase in width (centimeter), 22.683 increase ($\pm$ 2.021) in weight for every 1cm increase in length, and 13.670 increase ($\pm$ 3.854) in weight for every 1cm increase in width.