

# HW5

Emma Horton

2024-06-22

```
library(readr)
library(ggplot2)
```

Part 1

```
getwd()
```

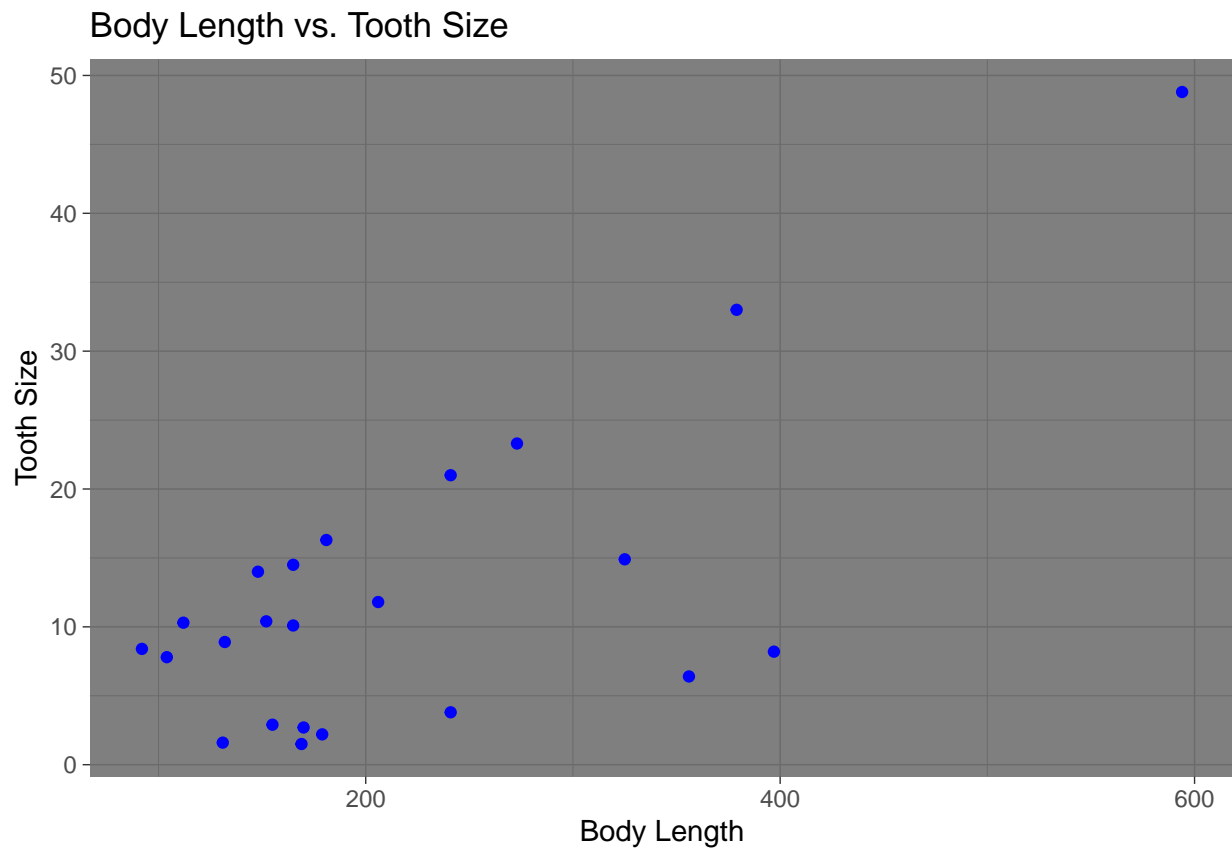
```
## [1] "/Users/emmahorton/DataScience/DataViz/Homework/HW5"
```

```
path <- file.path( "/Users/emmahorton/DataScience/DataViz/datasets/Sharks.csv")
sharks <- read.csv(path)
head(sharks)
```

```
##      Genus      Species BodyLength ToothSize Sex
## 1 Alopias   pelagicus      169        1.5   F
## 2 Alopias   pelagicus      179        2.2   F
## 3 Alopias   pelagicus      241        3.8   F
## 4 Alopias   pelagicus      170        2.7   M
## 5 Alopias superciliosus     325       14.9   M
## 6 Alopias   vulpinus      155        2.9   F
```

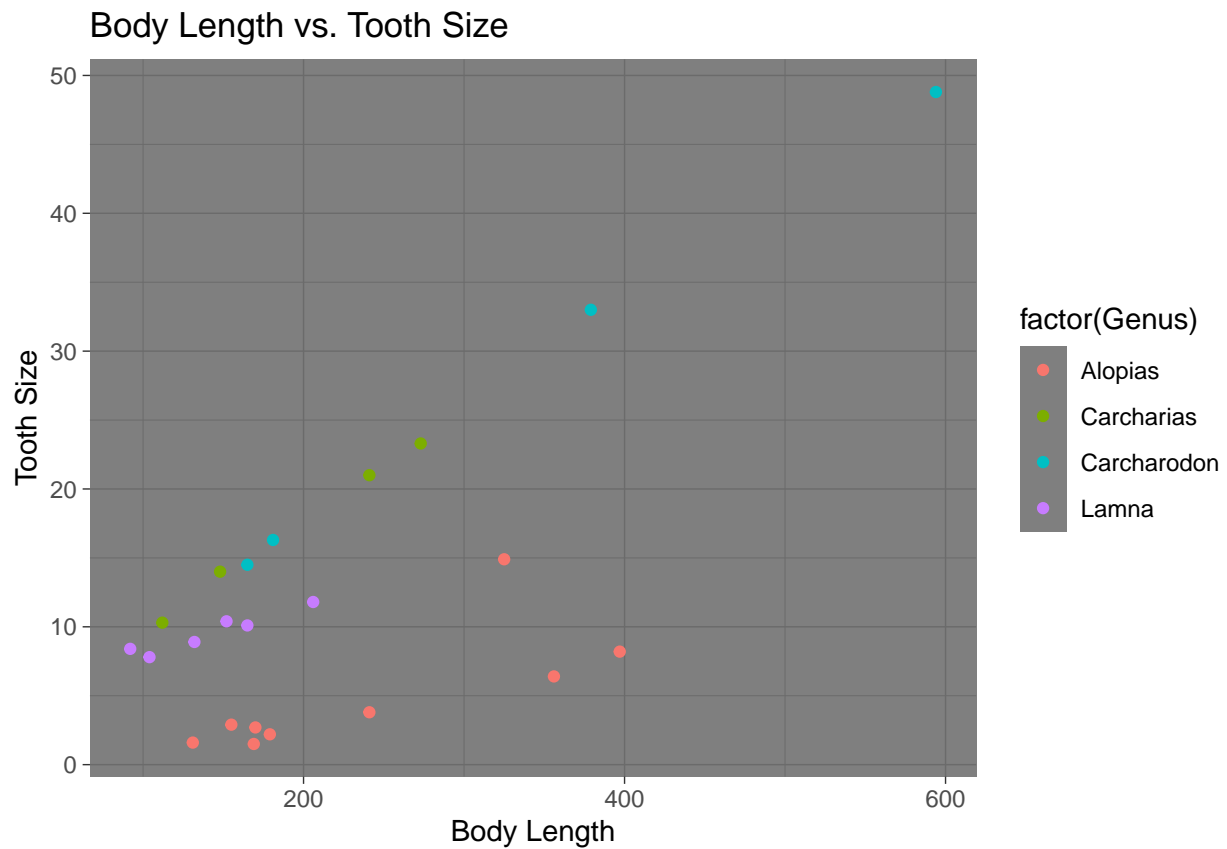
```
p <- ggplot(sharks, aes(BodyLength, ToothSize)) +
  geom_point(color = "blue") +
  labs(title = "Body Length vs. Tooth Size",
       x = "Body Length",
       y = "Tooth Size")

p + theme_dark()
```



Part 2

```
p <- ggplot(sharks, aes(BodyLength, ToothSize)) +  
  geom_point(aes(color = factor(Genus))) +  
  labs(title = "Body Length vs. Tooth Size",  
        x = "Body Length",  
        y = "Tooth Size")  
  
p + theme_dark()
```

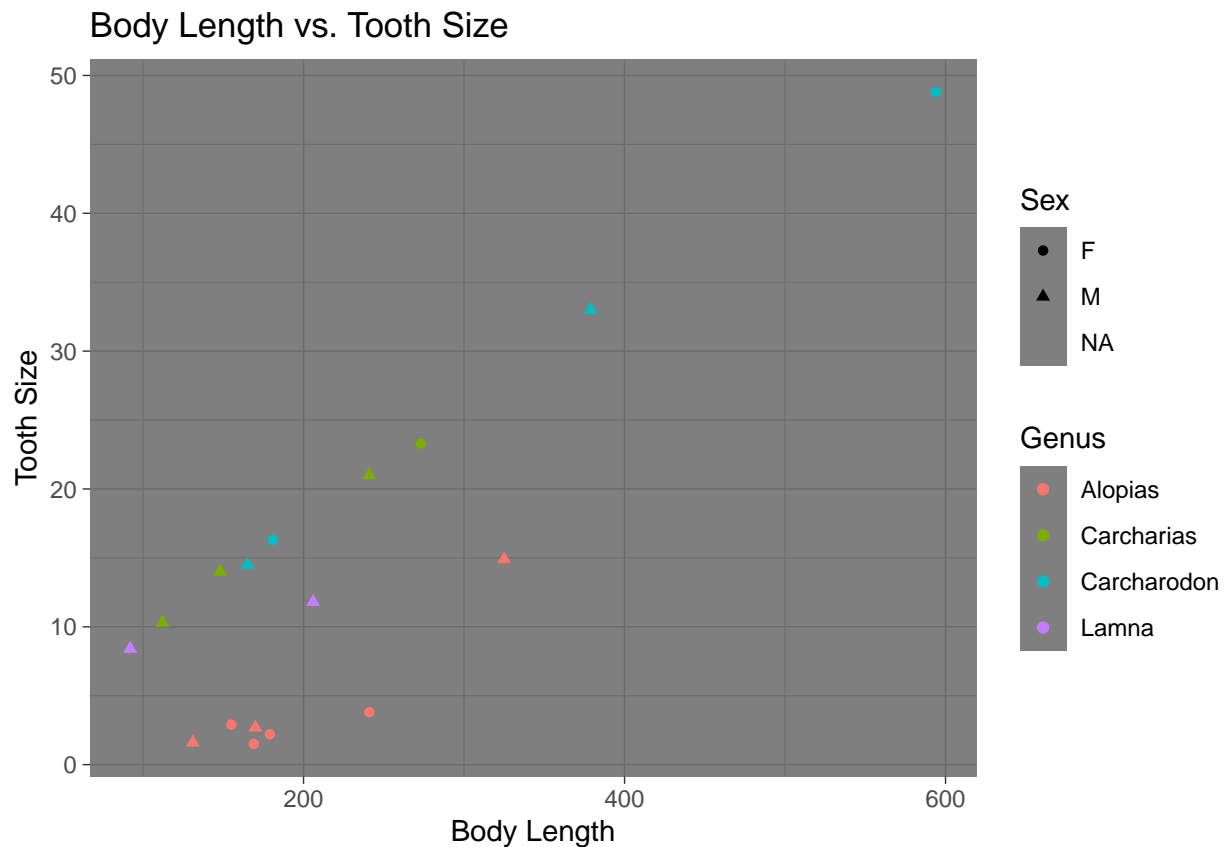


Part 3

```
p <- ggplot(sharks, aes(BodyLength, ToothSize)) +
  geom_point(aes(color = factor(Genus), shape = factor(Sex))) +
  labs(title = "Body Length vs. Tooth Size",
        x = "Body Length",
        y = "Tooth Size",
        color = "Genus",
        shape = "Sex")

p + theme_dark()
```

## Warning: Removed 6 rows containing missing values (`geom\_point()`).



```
p <- ggplot(sharks, aes(BodyLength, ToothSize)) +
  geom_point(aes(color = factor(Genus), shape = factor(Sex))) +
  labs(title = "Body Length vs. Tooth Size",
       x = "Body Length",
       y = "Tooth Size",
       color = "Genus",
       shape = "Sex")

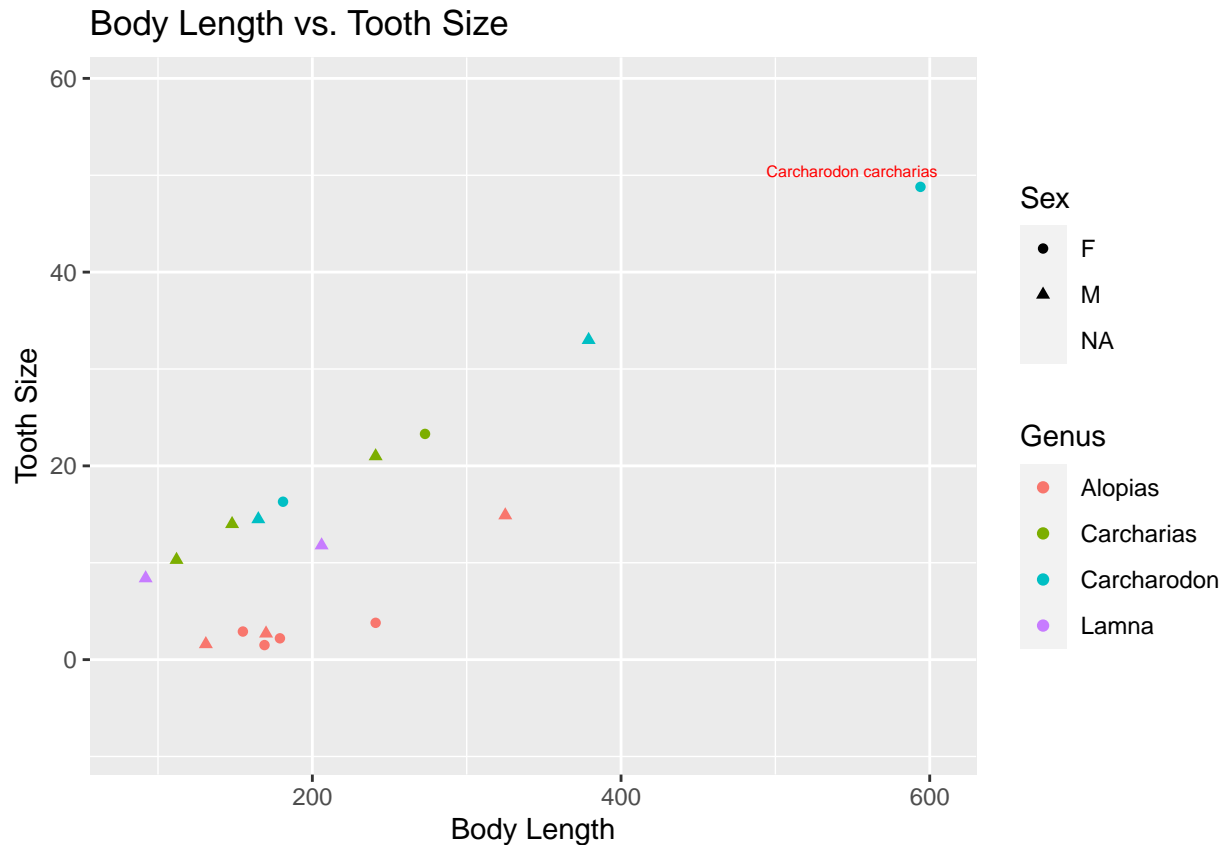
p <- p +
  xlim(min(sharks$BodyLength) - 10, max(sharks$BodyLength) + 10) +
  ylim(min(sharks$ToothSize) - 10, max(sharks$ToothSize) + 10)

BigGirlShark <- sharks[sharks$Sex == "F" & sharks$BodyLength > 550 & sharks$ToothSize > 40, ]

p <- p +
  geom_text(data = BigGirlShark, aes(label = paste(Genus, Species, sep = " "), vjust = -0.9, hjust = 0.9))

p

## Warning: Removed 6 rows containing missing values (`geom_point()`).
```

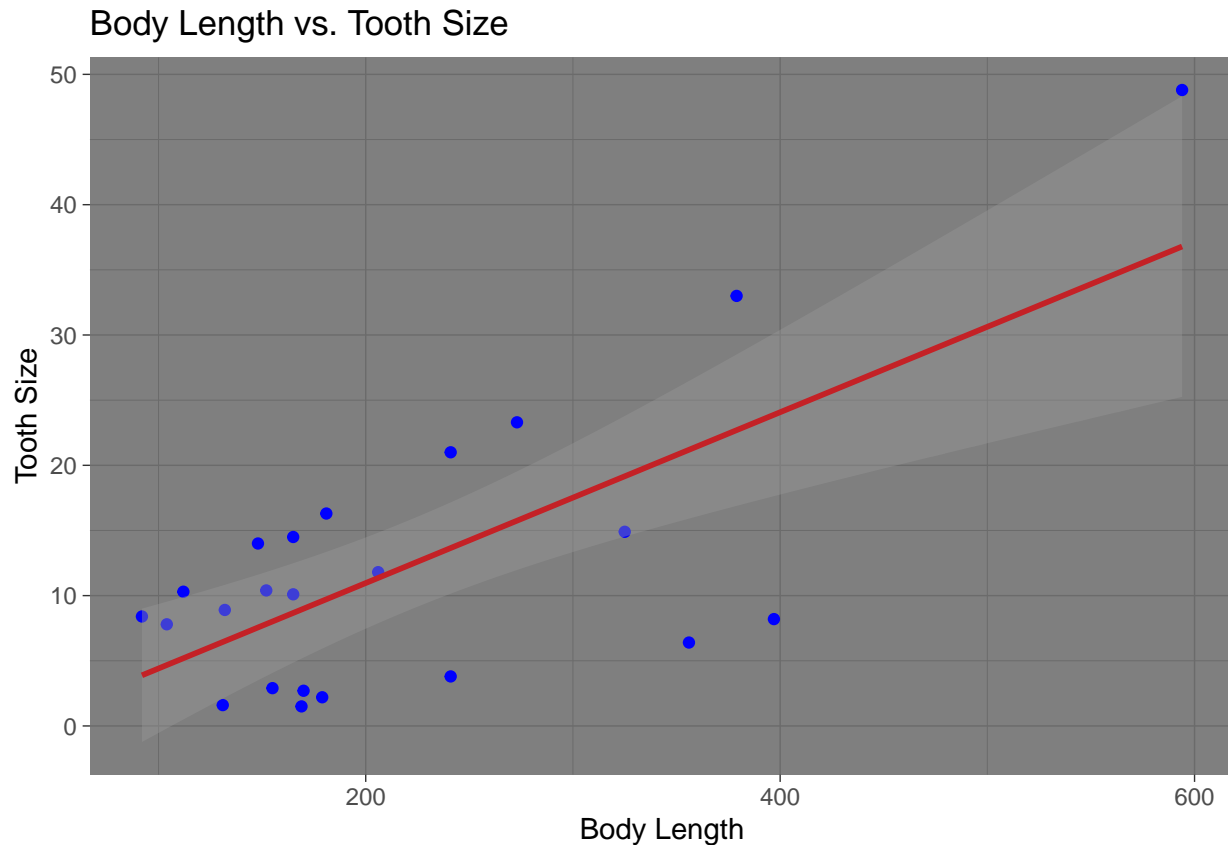


Part 5 The line does appear to capture the relationship generally but we see that there are a number of points that are not following the trend line likely due to genus of the shark.

```
p <- ggplot(sharks, aes(BodyLength, ToothSize)) +
  geom_point(color = "blue") +
  stat_smooth(method = "lm",
              col = "#C42126") +
  labs(title = "Body Length vs. Tooth Size",
       x = "Body Length",
       y = "Tooth Size")

p + theme_dark()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

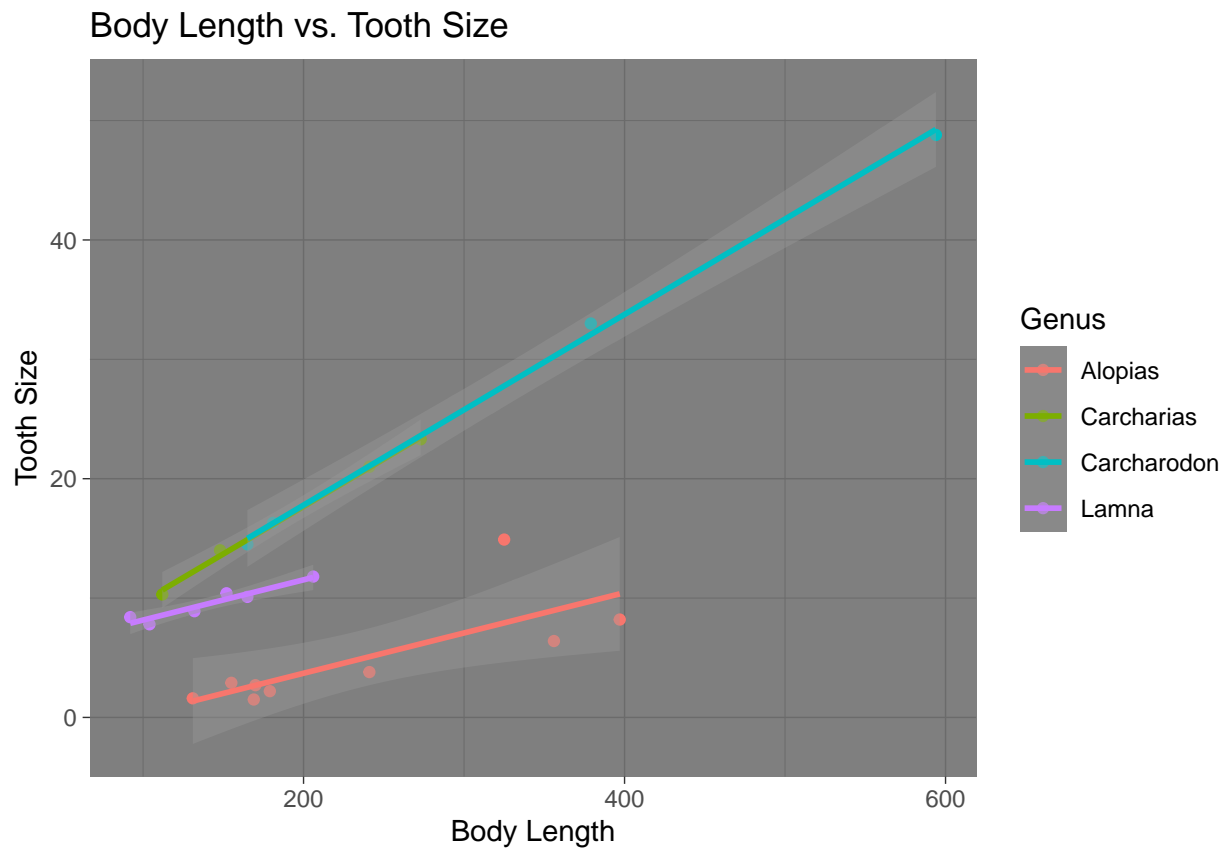


Part 6 We now see that more of the points fall within the standard error parameters because the tooth length is relative to the genus of the shark in addition to the size

```
p <- ggplot(sharks, aes(BodyLength, ToothSize, color = factor(Genus))) +
  geom_point() +
  stat_smooth(method = "lm") +
  labs(title = "Body Length vs. Tooth Size",
        x = "Body Length",
        y = "Tooth Size",
        color = "Genus") +
  theme_dark()
```

p

```
## `geom_smooth()` using formula = 'y ~ x'
```



```
sharks$Genus <- factor(sharks$Genus)

p <- ggplot(sharks, aes(BodyLength, ToothSize)) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(title = "Body Length vs. Tooth Size by Genus",
       x = "Body Length",
       y = "Tooth Size") +
  facet_wrap(~ Genus) +
  theme_dark()
```

p

```
## `geom_smooth()` using formula = 'y ~ x'
```

Body Length vs. Tooth Size by Genus

