Inclassmod3

```
library(palmerpenguins)
```

```
## Warning: package 'palmerpenguins' was built under R version 4.0.2
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.2
```

```
## -- Attaching packages ---------------------------------------------------------------------- tidyv
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.1     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

```
## Warning: package 'tidyr' was built under R version 4.0.2
```

```
## Warning: package 'readr' was built under R version 4.0.2
```

```
## Warning: package 'dplyr' was built under R version 4.0.2
```

```
## Warning: package 'stringr' was built under R version 4.0.2
```

```
## Warning: package 'forcats' was built under R version 4.0.2
```

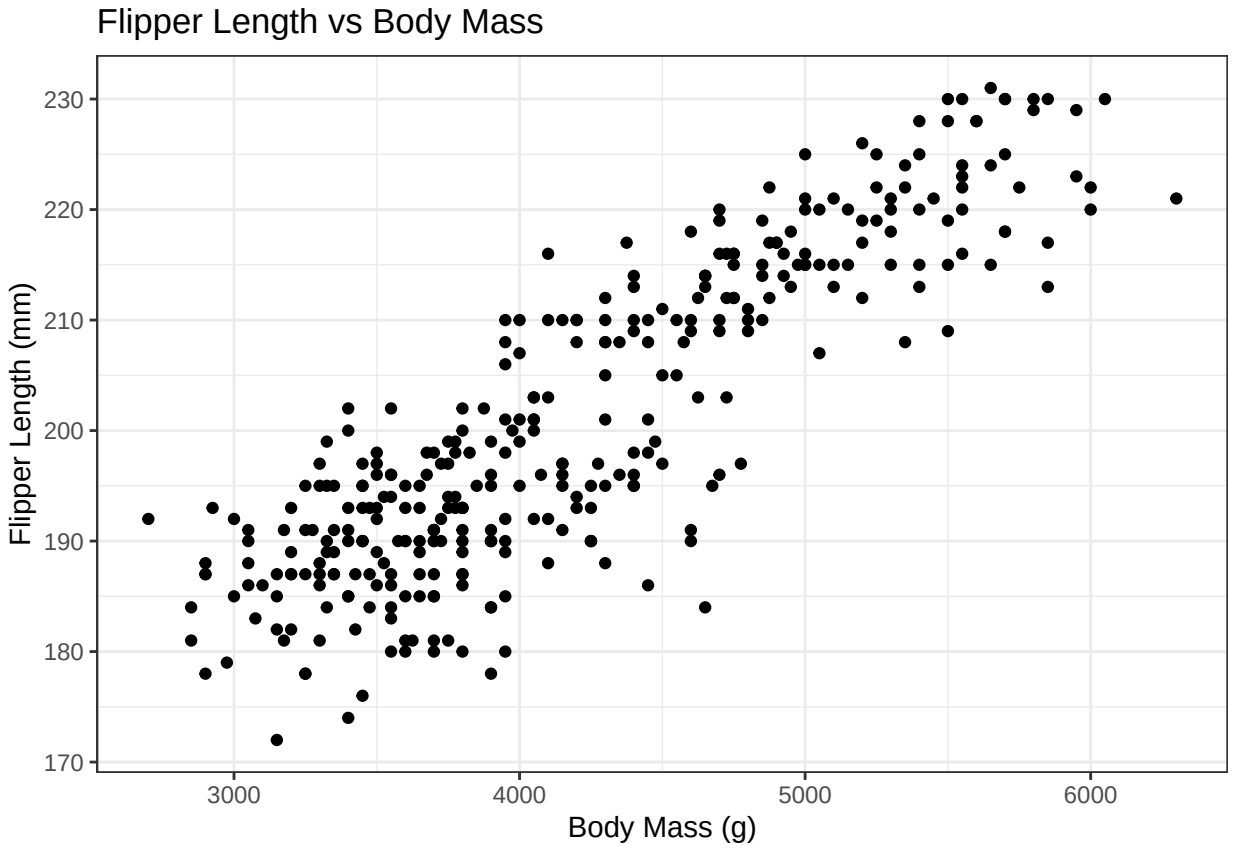
```
## -- Conflicts ------------------------------------------------------------------------- tidyverse_c
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

1

Produce a scatterplot of the two variables. How would you describe the relationship between the two variables? Be sure to label the axes and give an appropriate title. Based on the appearance of the plot, does a simple linear regression appear reasonable for the data?

```
penguins %>%
  ggplot(aes(body_mass_g, flipper_length_mm)) +
  geom_point() +
  labs(title="Flipper Length vs Body Mass", x="Body Mass (g)", y="Flipper Length (mm)") +
  theme_bw()
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```

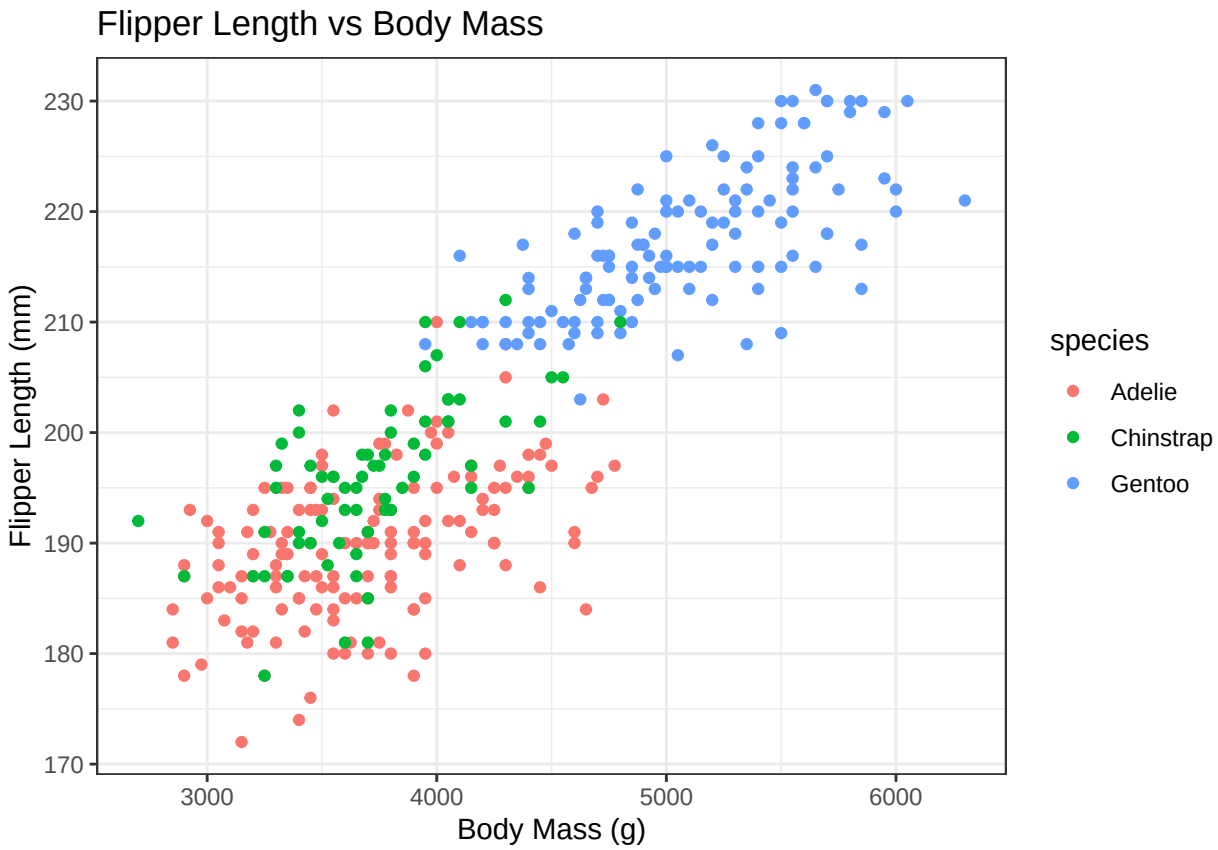


2

Produce a similar scatterplot, but with different colored plots for each species. How does this scatterplot influence your answer to the previous part?

```
penguins %>%
  ggplot(aes(body_mass_g, flipper_length_mm, color=species)) +
  geom_point() +
  labs(title="Flipper Length vs Body Mass", x="Body Mass (g)", y="Flipper Length (mm)") +
  theme_bw()
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```

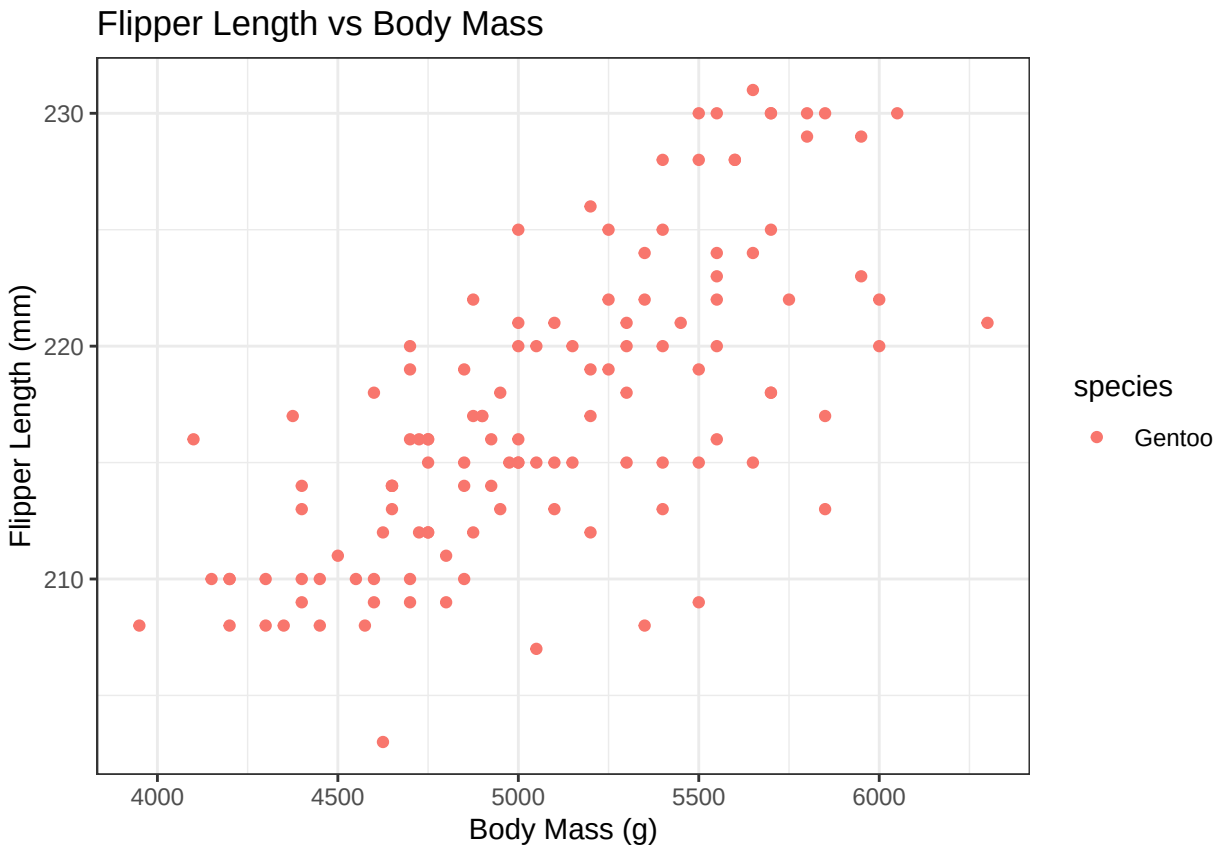


3

Regardless of your answer to the previous part, produce a scatterplot of body mass and flipper length for Gentoo penguins. Based on the appearance of the plot, does a simple linear regression appear reasonable for the data?

```
penguins %>%
  filter(species == "Gentoo") %>%
  ggplot(aes(body_mass_g, flipper_length_mm, color=species)) +
  geom_point() +
  labs(title="Flipper Length vs Body Mass", x="Body Mass (g)", y="Flipper Length (mm)") +
  theme_bw()
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



4

What is the correlation between body mass and flipper length for Gentoo penguins. Interpret this correlation contextually. How reliable is this interpretation?

```
cor(penguins[,c("flipper_length_mm", "body_mass_g")], use = "complete.obs")
```

```
##                   flipper_length_mm body_mass_g
## flipper_length_mm         1.0000000   0.8712018
## body_mass_g               0.8712018   1.0000000
```

5

Use the lm() function to fit a linear regression for body mass and flipper length for Gentoo penguins. Write out the estimated linear regression equation.

```
result<-lm(flipper_length_mm~body_mass_g, data=penguins)
summary(result)
```

```
##
## Call:
## lm(formula = flipper_length_mm ~ body_mass_g, data = penguins)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.7626  -4.9138   0.9891   5.1166  16.6392
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.367e+02  1.997e+00   68.47   <2e-16 ***
## body_mass_g 1.528e-02  4.668e-04   32.72   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.913 on 340 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.759,  Adjusted R-squared:  0.7583
## F-statistic:  1071 on 1 and 340 DF,  p-value: < 2.2e-16
```

$\hat{y} = 136.7 - 0.001528x$

6

Interpret the estimated slope contextually.

7

Does the estimated intercept make sense contextually?

8

Report the value of R^2 from this linear regression, and interpret its value contextually.