

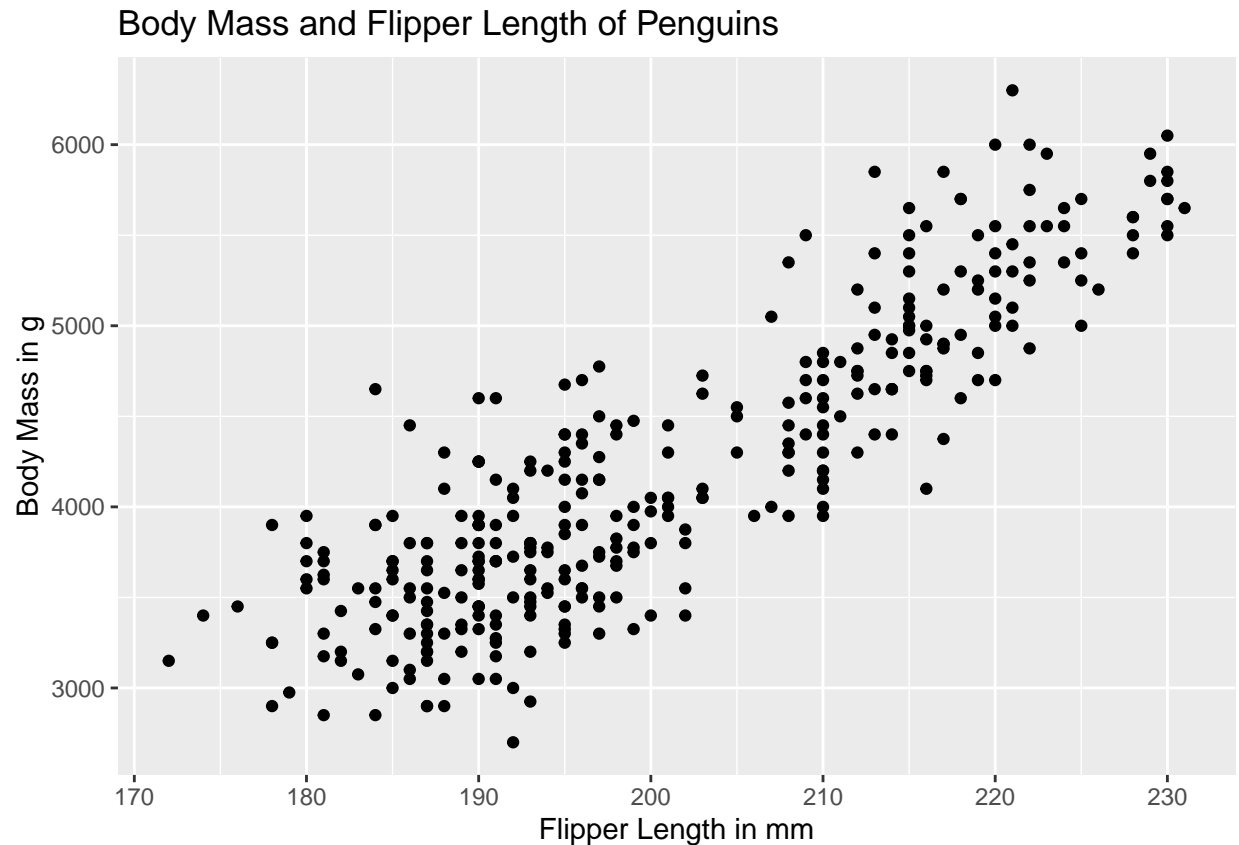
Guided Question Set 3 Solutions

```
library(tidyverse)
library(palmerpenguins)
Data<-penguins
```

1)

```
##create scatterplot
ggplot(Data, aes(x=flipper_length_mm,y=body_mass_g))+
  geom_point()+
  labs(x="Flipper Length in mm",
       y="Body Mass in g",
       title="Body Mass and Flipper Length of Penguins")
```

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

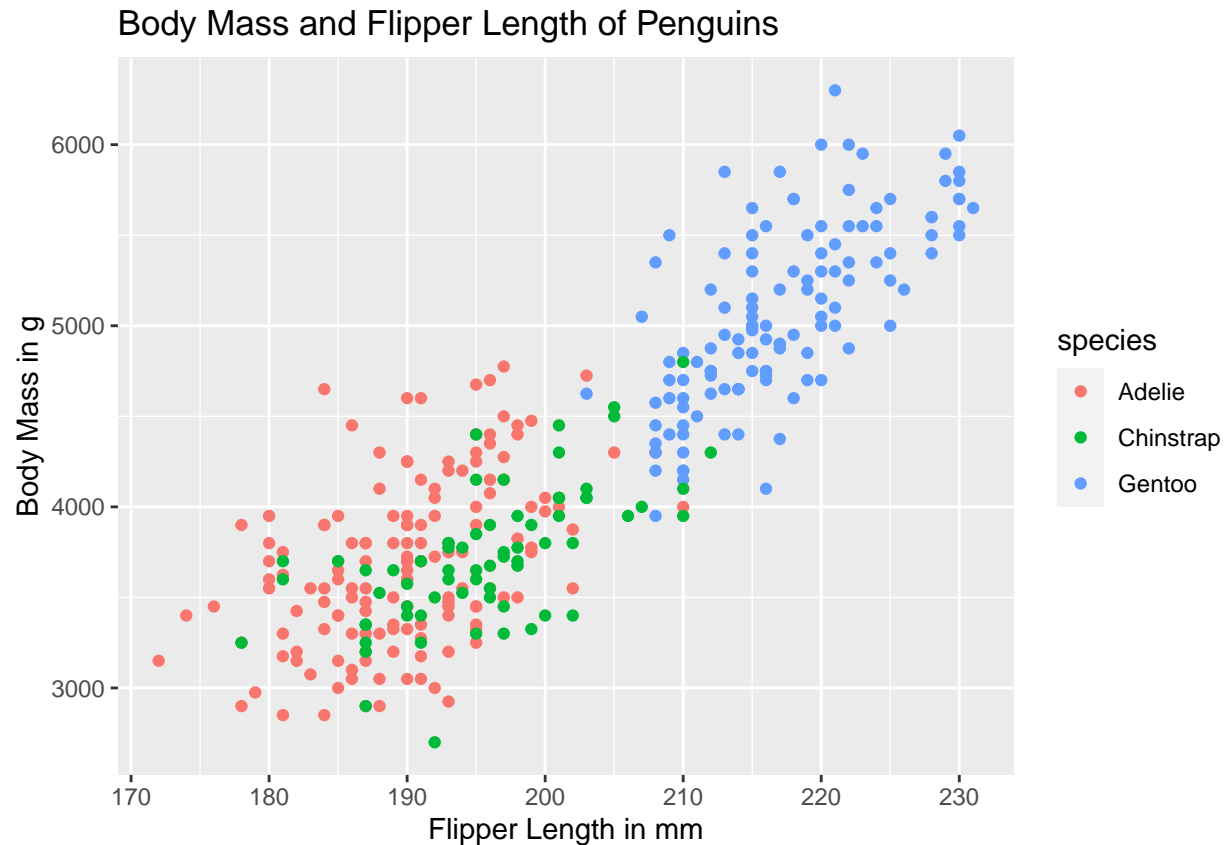


The scatterplot of body mass and flipper length of the penguins is displayed above. A simple linear regression appears reasonable for the data, as we have a strong positive linear relationship between the variables. Generally, as flipper length increases, the body mass increases.

2)

```
ggplot(Data, aes(x=flipper_length_mm,y=body_mass_g, color=species))+  
  geom_point()+  
  labs(x="Flipper Length in mm",  
        y="Body Mass in g",  
        title="Body Mass and Flipper Length of Penguins")
```

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



Within each species, we still see a positive linear association between body mass and flipper length. However, Gentoo penguins appear to be larger than Adelie and Chinstrap penguins. So it may be worth to consider separate regressions for each species (or at least separate Gentoos from the Adelies and Chinstraps).

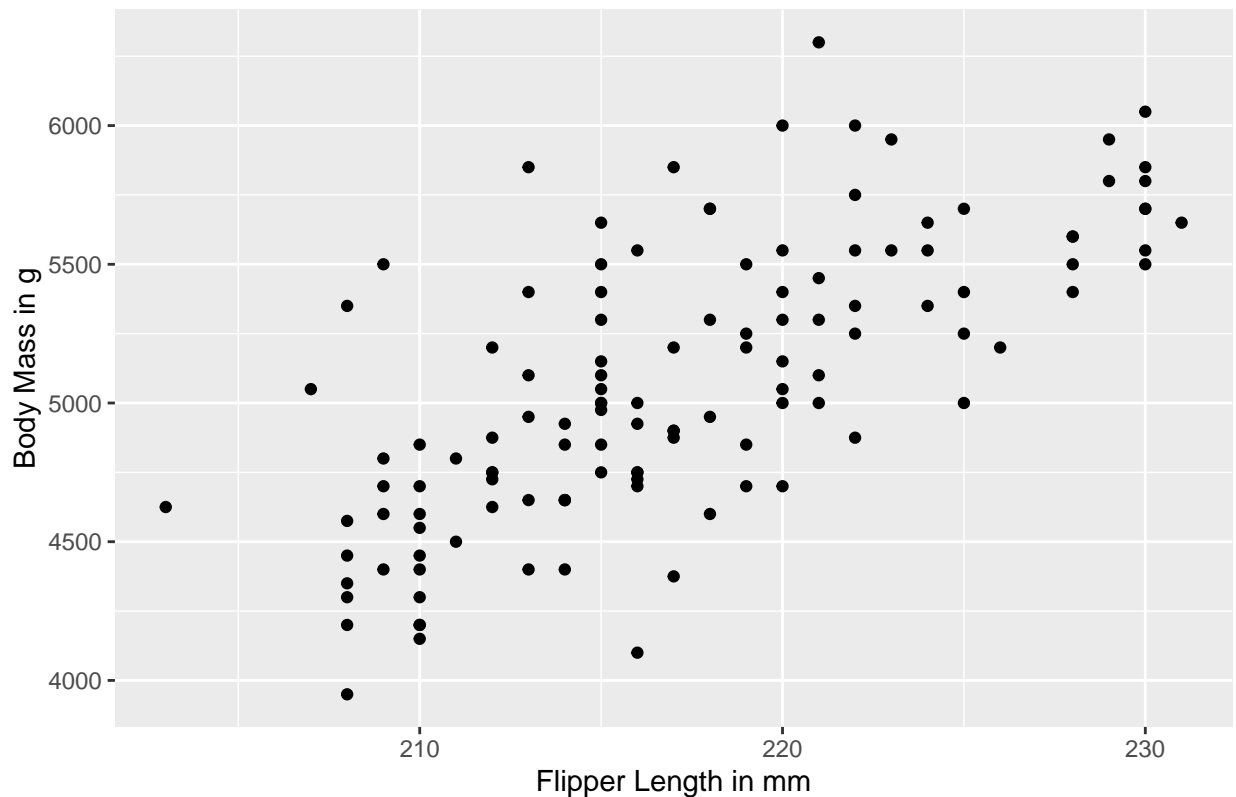
3)

```
gentoo<-Data %>%
  filter(species=="Gentoo")

ggplot(gentoo, aes(x=flipper_length_mm,y=body_mass_g))+
  geom_point()+
  labs(x="Flipper Length in mm",
       y="Body Mass in g",
       title="Body Mass and Flipper Length of Gentoo Penguins")
```

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

Body Mass and Flipper Length of Gentoo Penguins



For Gentoo penguins, there appears to be a positive linear association between body mass and flipper length.

4)

```
cor(gentoo$flipper_length_mm, gentoo$body_mass_g, use = "complete.obs")
```

```
## [1] 0.7026665
```

The correlation is about 0.7027. This value indicates a reasonably strong positive linear association between body mass and flipper length. Since the scatterplot shows a linear association is reasonable, we can interpret the correlation reliably.

5)

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

```
##
## Call:
## lm(formula = body_mass_g ~ flipper_length_mm, data = gentoo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -911.18 -235.76  -51.93   170.75 1015.71
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6787.281    1092.552   -6.212 7.65e-09 ***
## flipper_length_mm    54.623      5.028   10.863 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 360.2 on 121 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.4937, Adjusted R-squared:  0.4896
## F-statistic: 118 on 1 and 121 DF, p-value: < 2.2e-16
```

The estimated regression equation is $\hat{y} = -6787.281 + 54.623x$.

6)

For Gentoo penguins, the predicted body mass increases by 54.623g for each additional mm increase in flipper length.

7)

The estimated intercept of -6787.281 does not make sense contextually, as a penguin cannot have a negative value for body mass, neither can a penguin have a flipper length of 0mm. This is a by product of extrapolation.

8)

The R^2 is 0.4937. This means that about 49.37% of the variation in body mass can be explained by flipper length, for Gentoo penguins.

9)

The estimated value for the standard deviation of the error terms is the residual standard error, 360.2.

10)

```
yhat<-result$coefficients[1] + result$coefficients[2]*220
yhat
```

```
## (Intercept)
##      5229.67
```

The predicted body mass of a Gentoo penguin with flipper length 220mm is 5229.67g.

11)

```
##create ANOVA table
anova.tab<-anova(result)
anova.tab
```

```
## Analysis of Variance Table
##
## Response: body_mass_g
##              Df    Sum Sq  Mean Sq F value    Pr(>F)
## flipper_length_mm    1 15308045 15308045  118.01 < 2.2e-16 ***
## Residuals          121 15696203   129721
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##R2
SST<-sum(anova.tab$"Sum Sq")
SST
```

```
## [1] 31004248
```

```
anova.tab$"Sum Sq"[1]/SST
```

```
## [1] 0.4937402
```

$$R^2 = \frac{SS_R}{SS_T} = \frac{15308045}{31004248} = 0.4937402.$$

12)

$$H_0 : \beta_1 = 0, H_a : \beta_1 \neq 0$$

13)

$$F = \frac{MS_R}{MS_{res}} = \frac{15308045}{129721} = 118.01.$$

14)

The p-value is less than 0.05, so we reject the null hypothesis. Our data support the claim that there is a linear relationship between body mass and flipper length for Gentoo penguins.