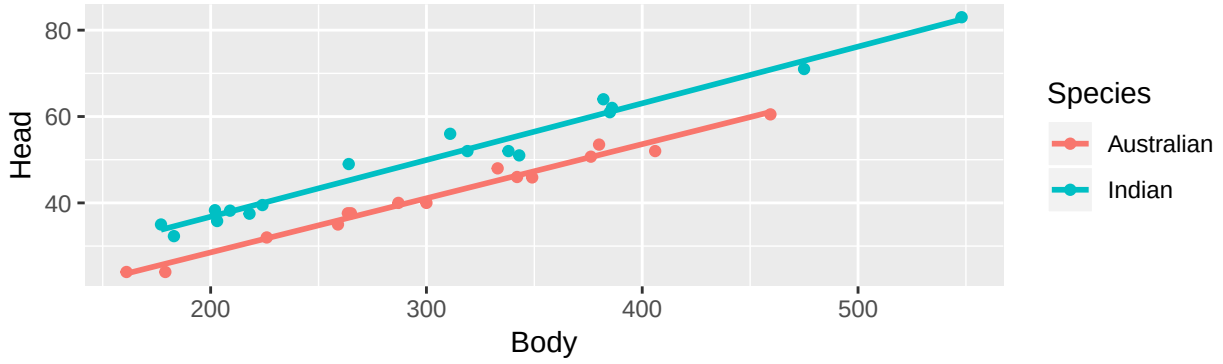
Two Lines: Crocodiles!!

```
head(crocs)
```

```
##    Body Head    Species
## 1 338.0 52.0     Indian
## 2 333.0 48.0 Australian
## 3 202.0 38.3     Indian
## 4 406.0 52.0 Australian
## 5 459.4 60.5 Australian
## 6 264.0 49.0     Indian
```

```
ggplot(data = crocs, mapping = aes(x = Body, y = Head, color = Species)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE)
```



2 lines by filtering to create separate data sets

```
aus_crocs <- crocs %>% filter(Species == "Australian")
aus_fit <- lm(Head ~ Body, data = aus_crocs)
summary(aus_fit)
```

```
##
## Call:
## lm(formula = Head ~ Body, data = aus_crocs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3529 -0.9968  0.0824  0.7419  2.7973
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.463022   1.523732   2.273   0.0407 *
## Body        0.125344   0.004819  26.010 1.35e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.504 on 13 degrees of freedom
## Multiple R-squared:  0.9811, Adjusted R-squared:  0.9797
## F-statistic: 676.5 on 1 and 13 DF,  p-value: 1.35e-12
```

```r
ind_crocs <- crocs %>% filter(Species == "Indian")
ind_fit <- lm(Head ~ Body, data = ind_crocs)
summary(ind_fit)
```

```
##
## Call:
## lm(formula = Head ~ Body, data = ind_crocs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5756 -1.6627 -0.0904  1.2208  4.6261
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.538438   1.861787    5.66 4.53e-05 ***
## Body         0.131304   0.005791   22.68 5.08e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.503 on 15 degrees of freedom
## Multiple R-squared:  0.9717, Adjusted R-squared:  0.9698
## F-statistic: 514.2 on 1 and 15 DF,  p-value: 5.08e-13
```

Questions we'd like to be able to answer (but can't with this output):

1. Is there statistically significant evidence that the intercepts for these lines are different?
2. Is there statistically significant evidence that the slopes for these lines are different?

2 parallel lines (same slope)

- Our Goal: Equations for two lines

$$\text{Predicted Head Length for Australian Crocs} = \hat{\beta}_0^{Australian} + \hat{\beta}_1 \times (\text{Body Length})$$
$$\text{Predicted Head Length for Indian Crocs} = \hat{\beta}_0^{Indian} + \hat{\beta}_1 \times (\text{Body Length})$$

- Note: Different intercepts, same slope.

```
parallel_lines_fit <- lm(Head ~ Body + Species, data = crocs)
summary(parallel_lines_fit)
```

```
##
## Call:
## lm(formula = Head ~ Body + Species, data = crocs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4959 -1.4218 -0.0842  1.0117  4.6405
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.265418   1.309167    1.73   0.0942 .
## Body          0.129261   0.003904   33.11  < 2e-16 ***
## SpeciesIndian 8.893772   0.737538   12.06 8.05e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.082 on 29 degrees of freedom
## Multiple R-squared:  0.977,  Adjusted R-squared:  0.9755
## F-statistic:   617 on 2 and 29 DF,  p-value: < 2.2e-16
```

- R gives us a single combined equation:

$$\text{Predicted Head Length} = \hat{\beta}_0 + \hat{\beta}_1(\text{Body Length}) + \hat{\beta}_2\text{SpeciesIndian}$$

$$\text{Predicted Head Length} = 2.27 + 0.13(\text{Body Length}) + 8.89\text{SpeciesIndian}$$

What is the `SpeciesIndian` variable?

- Behind the scenes, R creates a new **indicator variable** called `SpeciesIndian`:

$$\texttt{SpeciesIndian} = \begin{cases} 1 & \text{if the species for crocodile } i \text{ is Indian.} \\ 0 & \text{otherwise (in this case, the species is Australian)} \end{cases}$$

- R doesn't modify the data frame (it creates a secret copy in the background), but it would look like this:

```
head(crocs)
```

```
##    Body Head    Species SpeciesIndian
## 1 338.0 52.0     Indian             1
## 2 333.0 48.0 Australian             0
## 3 202.0 38.3     Indian             1
## 4 406.0 52.0 Australian             0
## 5 459.4 60.5 Australian             0
## 6 264.0 49.0     Indian             1
```

Above, we obtained this estimated equation:

$$\text{Predicted Head Length} = 2.27 + 0.13(\text{Body Length}) + 8.89\text{SpeciesIndian}$$

What is the estimated equation describing the relationship between body length and head length, for Australian crocodiles?

What is the estimated equation describing the relationship between body length and head length, for Indian crocodiles?

What is the interpretation of $\widehat{\beta}_0 = 2.27$?

What is the interpretation of $\widehat{\beta}_1 = 0.13$?

What is the interpretation of $\widehat{\beta}_2 = 8.89$?

Using the output from the summary function, conduct a test of the claim that the intercept of the line describing the relationship between body length and head length in the population of all Australian crocodiles is the same as the intercept of the line describing the relationship between body length and head length in the population of all Indian crocodiles.

Lines with different slopes (interactions)

To get different slopes, use `Body * Species` instead of `Body + Species`:

```
two_lines_fit <- lm(Head ~ Species * Body, data = crocs)
summary(two_lines_fit)
```

```
##
## Call:
## lm(formula = Head ~ Species * Body, data = crocs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5756 -1.3294 -0.0040  0.9646  4.6261
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        3.463022   2.126572   1.628   0.1146
## SpeciesIndian      7.075415   2.638253   2.682   0.0121 *
## Body               0.125344   0.006726  18.637   <2e-16 ***
## SpeciesIndian:Body 0.005959   0.008296   0.718   0.4785
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.099 on 28 degrees of freedom
## Multiple R-squared:  0.9775, Adjusted R-squared:  0.975
## F-statistic: 404.6 on 3 and 28 DF,  p-value: < 2.2e-16
```

What is the estimated equation from this model?

What is the estimated equation describing the relationship between body length and head length, for Australian crocodiles?

What is the estimated equation describing the relationship between body length and head length, for Indian crocodiles?

What is the interpretation of $\widehat{\beta}_0 = 3.463$?

What is the interpretation of $\widehat{\beta}_1 = 7.075$?

What is the interpretation of $\widehat{\beta}_2 = 0.125$?

What is the interpretation of $\widehat{\beta}_3 = 0.006$?

Using the output from the summary function, conduct a test of the claim that the *slope* of the line describing the relationship between body length and head length in the population of all Australian crocodiles is the same as the *slope* of the line describing the relationship between body length and head length in the population of all Indian crocodiles.