UsingModels2.R

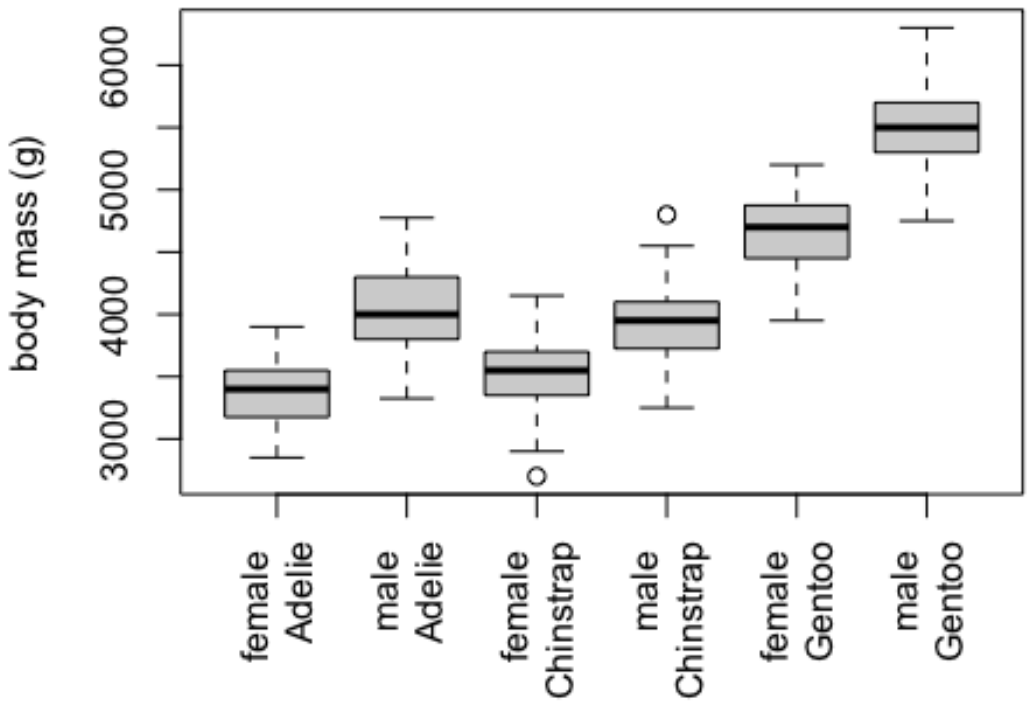
Feipeng Huang

2022-12-01

```
require(palmerpenguins)

## Loading required package: palmerpenguins

boxplot(body_mass_g ~ sex*species, data = penguins, las = 3, names =
c("female \n Adelie", "male \n Adelie", "female \n Chinstrap", "male \n
Chinstrap", "female \n Gentoo", "male \n Gentoo"), ylab = "body mass (g)",
xlab = "")
```




```
#2 Based on the boxplots, I think male penguins are significantly heavier
than female penguins of the same species and the difference is significant
because the box for male is higher than the box for female of the same
species and the boxes do not overlap much.
#3 I think adding sex to a model that already includes species will improve
the model fit because there are significant variations in sex that we need to
```

```
capture.
#4
fit_both = lm(body_mass_g ~ sex * species, data = penguins)
summary(fit_both)

##
## Call:
## lm(formula = body_mass_g ~ sex * species, data = penguins)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -827.21 -213.97   11.03  206.51  861.03
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                3368.84      36.21  93.030  < 2e-16 ***
## sexmale                     674.66      51.21  13.174  < 2e-16 ***
## speciesChinstrap            158.37      64.24   2.465  0.01420 *
## speciesGentoo              1310.91      54.42  24.088  < 2e-16 ***
## sexmale:speciesChinstrap  -262.89      90.85  -2.894  0.00406 **
## sexmale:speciesGentoo      130.44      76.44   1.706  0.08886 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 309.4 on 327 degrees of freedom
##   (11 observations deleted due to missingness)
## Multiple R-squared:  0.8546, Adjusted R-squared:  0.8524
## F-statistic: 384.3 on 5 and 327 DF,  p-value: < 2.2e-16

#5 female Adelie
#6 Intercept, speciesChinstrap
#7 3527.206
summary(fit_both)$coefficient [1 , 1] + summary(fit_both)$coefficient [3 , 1]

## [1] 3527.206

#8 3527.206
Chinstrap = subset(penguins, species == "Chinstrap")
Chinstrap_female = subset(Chinstrap, sex == "female")
mean(Chinstrap_female$body_mass_g)

## [1] 3527.206

aggregate(body_mass_g ~ species*sex, data = penguins, FUN = mean)

##     species    sex body_mass_g
## 1    Adelie female    3368.836
## 2 Chinstrap female    3527.206
## 3    Gentoo female    4679.741
## 4    Adelie   male    4043.493
```

```
## 5 Chinstrap   male    3938.971
## 6    Gentoo   male    5484.836
```