

Penguin Assignment

Anonymous

06/12/2022

Organising and cleaning data:

Setting working directory as “PenguinProjects” folder (please adjust to suit your working directory):

```
setwd("~/RWorkingDirectory/PenguinProjects")
```

Loading in libraries, data and functions:

```
source("functions/libraries.r")
source("functions/cleaning.r")
source("functions/plotting.r")
```

Creating a new folder for the raw data and saving the raw data:

```
#Creating a new data folder
dir.create("data")

#Creating a new raw data folder
dir.create("data/data_raw")

#Saving the raw data
write.csv(penguins_raw, "data/data_raw/penguins_raw.csv")
```

Cleaning the data, creating a new folder for the clean data and saving the clean data:

```
#Using function from "cleaning.r" to clean the data
penguins_clean <- cleaning(penguins_raw)

#Creating a new clean data folder
dir.create("data/data_clean")

#Saving the clean data
write.csv(penguins_clean, "data/data_clean/penguins_clean_06-12-2022.csv")
```

Looking at the data:

```
head(penguins_clean)
```

Preparing data for analysis:

Removing rows without sex or body mass recorded:

```
#Using function from "cleaning.r" to prepare data
penguins_mass <- remove_empty_body_mass(penguins_clean)

#Defining factors and numerical variables
penguins_mass$species <- as.factor(penguins_mass$species)
penguins_mass$sex <- as.factor(penguins_mass$sex)
penguins_mass$body_mass_g <- as.numeric(penguins_mass$body_mass_g)

#Saving mass data
write.csv(penguins_mass, "data/data_clean/penguins_mass_06-12-2022.csv")
```

Looking at the data:

```
head(penguins_mass, 3)
```

```
## # A tibble: 3 x 3
##   species                sex  body_mass_g
##   <fct>                <fct>      <dbl>
## 1 Adelie Penguin (Pygoscelis adeliae) MALE      3750
## 2 Adelie Penguin (Pygoscelis adeliae) FEMALE     3800
## 3 Adelie Penguin (Pygoscelis adeliae) FEMALE     3250
```

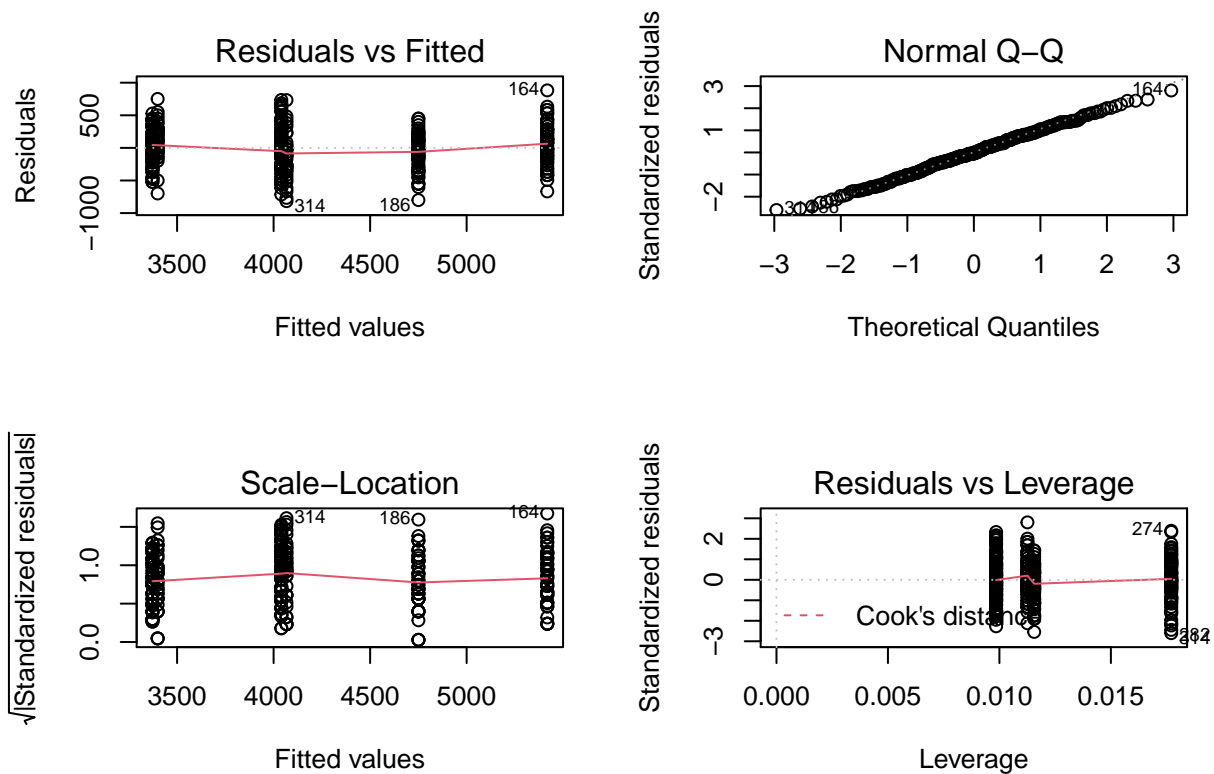
Analysing the data:

Q: To what extent does the species and sex of a penguin influence body mass?

Creating a model with no interaction and testing assumptions of statistical analysis:

```
#Creating the model
lmSpecies_Sex_Mass_No_Interaction <- aov(body_mass_g ~ species + sex, data = penguins_mass)

#Plotting to test assumptions
par(mfrow = c(2,2))
plot(lmSpecies_Sex_Mass_No_Interaction)
```



Overall, the data appears to meet assumptions of ANOVA.

Running ANOVA:

```
summary(lmSpecies_Sex_Mass_No_Interaction)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
species	2	145190219	72595110	724.2	<2e-16 ***
sex	1	37090262	37090262	370.0	<2e-16 ***
Residuals	329	32979185	100241		

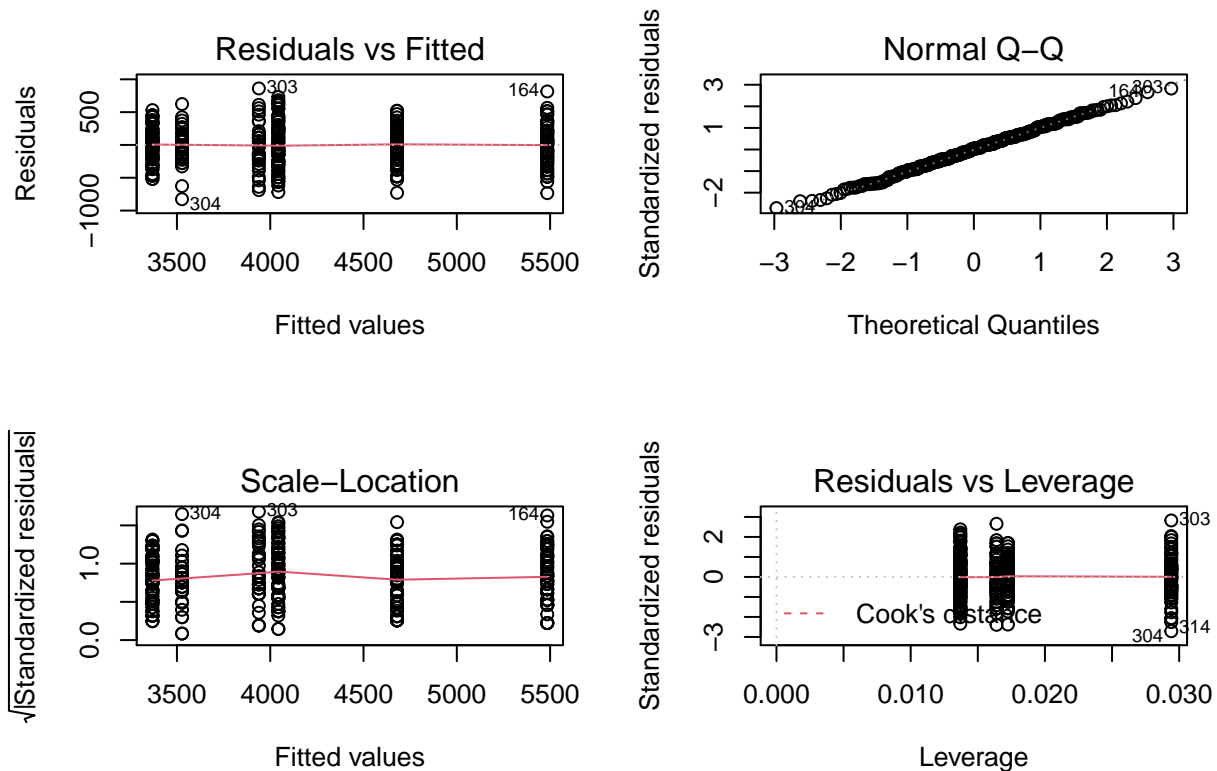
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Both species and sex appear to have a significant effect on body mass.

Building a model that takes into account interaction between species and sex:

```
#Creating the model
lmSpecies_Sex_Mass_Interaction <- aov(body_mass_g ~ species * sex, data = penguins_mass)

#Plotting to test assumptions
par(mfrow = c(2,2))
plot(lmSpecies_Sex_Mass_Interaction)
```



Overall, the data appears to fit the assumptions of ANCOVA.

Running ANCOVA:

```
summary(lmSpecies_Sex_Mass_Interaction)
```

	##	Df	Sum Sq	Mean Sq	F value	Pr(>F)
species	## species	2	145190219	72595110	758.358	< 2e-16 ***
sex	## sex	1	37090262	37090262	387.460	< 2e-16 ***
species:sex	## species:sex	2	1676557	838278	8.757	0.000197 ***
Residuals	## Residuals	327	31302628	95727		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

There is a significant effect of interaction between species and sex on body mass.

Comparing the two models:

```
anova(lmSpecies_Sex_Mass_No_Interaction, lmSpecies_Sex_Mass_Interaction)
```

```
## Analysis of Variance Table
##
## Model 1: body_mass_g ~ species + sex
## Model 2: body_mass_g ~ species * sex
##   Res.Df    RSS Df Sum of Sq   F    Pr(>F)
## 1      329 32979185
## 2      327 31302628  2   1676557 8.757 0.0001973 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Adding an interaction between species and sex significantly improves the model.

Running a Tukey HSD test:

```
#Using TukeyHSD test to compare between groups (results hidden here to keep document tidy)
TukeyHSD(lmSpecies_Sex_Mass_Interaction)
```

The Tukey HSD test suggests significant differences between body mass of penguins grouped by species and sex.

The comparisons between groups which are not significantly different are:

- 1) Chinstrap females and Adelie females
- 2) Chinstrap males and Adelie males

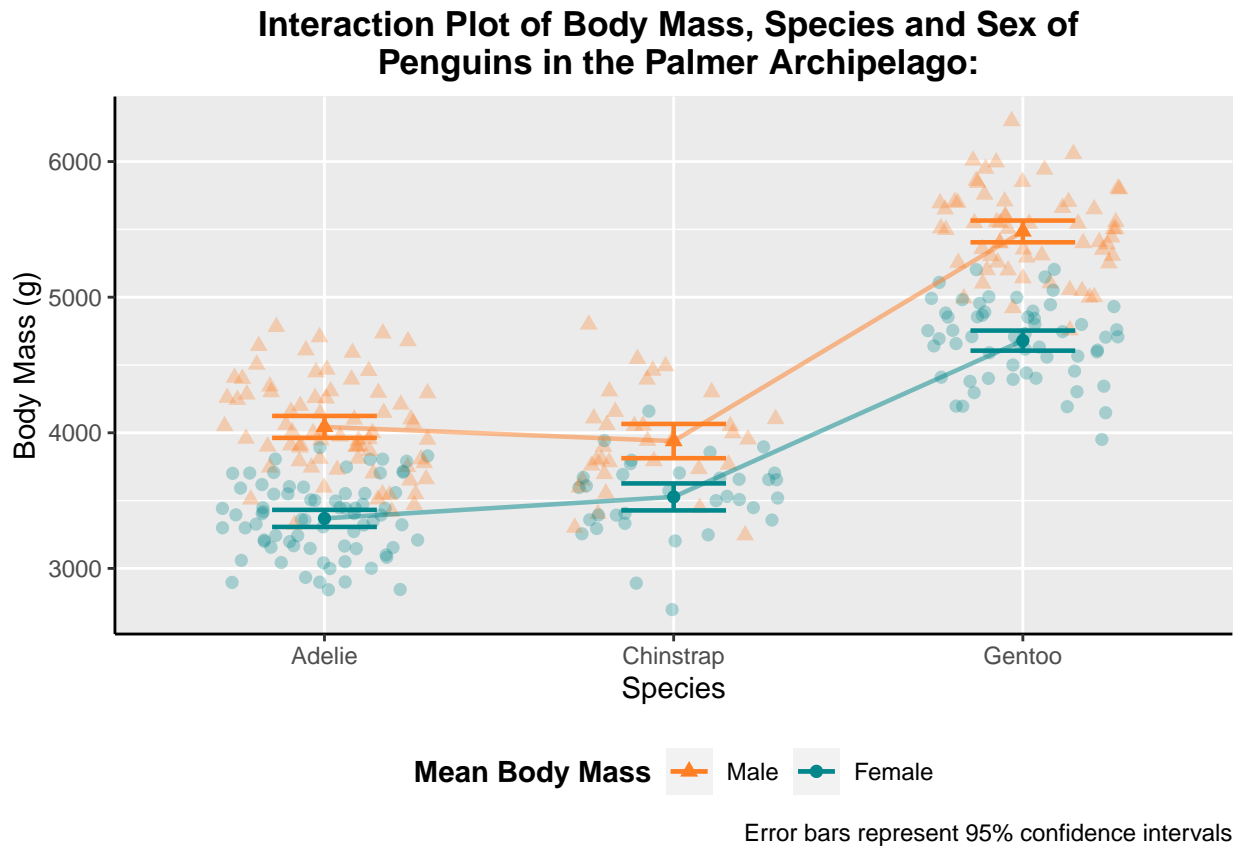
Producing figures to display the data:

Creating a new folder to store figures in:

```
dir.create("figures")
```

Producing a dot plot with lines for each species:

```
#Using function from "plotting.R" to produce graph  
plot_mass_figure(penguins_mass)
```



Saving the figure as a PNG file:

```
#Using function from "plotting.R" to save figure as PNG  
save_mass_plot_png(penguins_mass, "figures/fig_01_06-12-2022.png", 20, 600, 1.4)
```

Saving the figure as an SVG file:

```
#Using function from "plotting.R" to save figure as SVG  
save_mass_plot_svg(penguins_mass, "figures/fig_01_06-12-2022.svg", 20, 1.4)
```