Palmer Penguins Project

Olumide Olaoye

11/14/2021

Palmer Penguins Project: Practice Analysis

Setting up my environment

Notes: setting up my R environment by installing and loading the 'tidyverse', 'ggplot2' and 'palmerpenguins' packages

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.0     v forcats 0.5.1
```

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

```
library(ggplot2)
library(palmerpenguins)
```

Structure of the penguins dataset

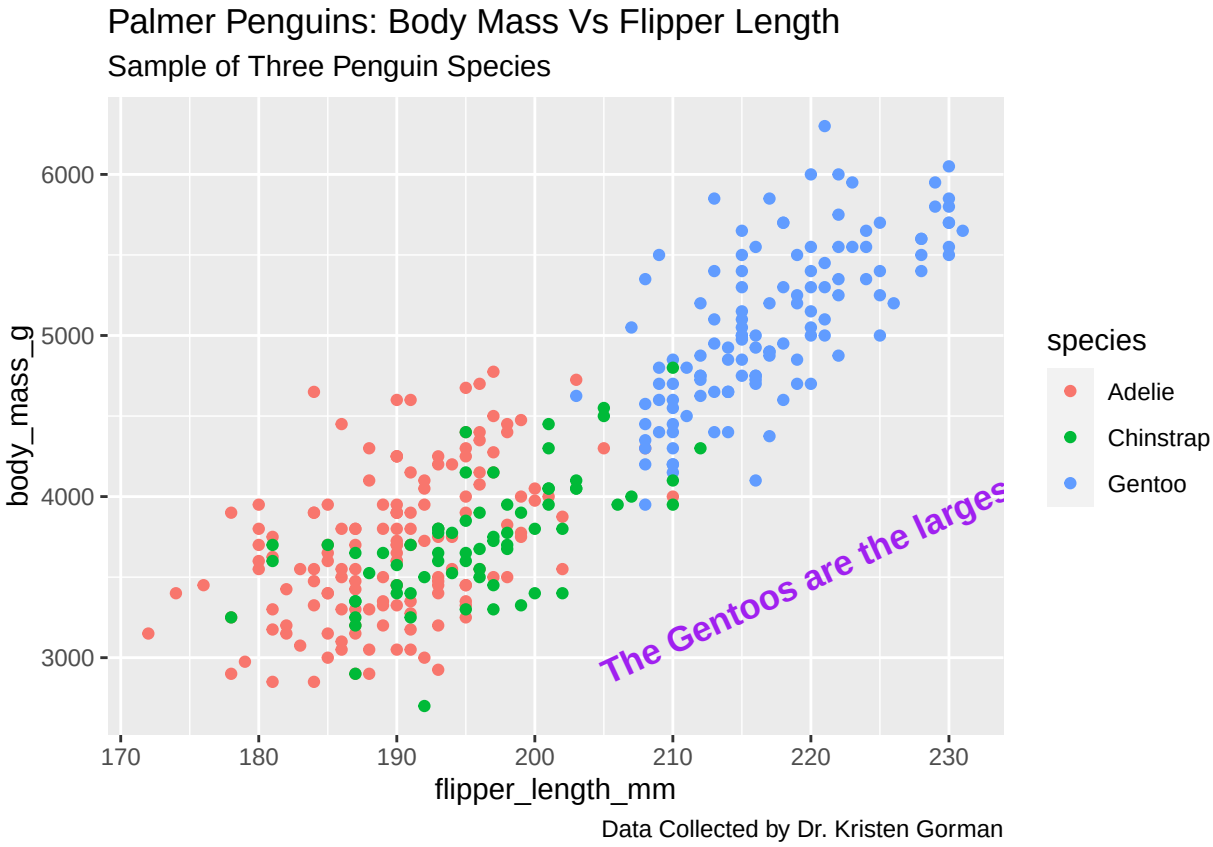
```
str(penguins)
```

```
## tibble [344 x 8] (S3: tbl_df/tbl/data.frame)
##  $ species          : Factor w/ 3 levels "Adelie","Chinstrap",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ island           : Factor w/ 3 levels "Biscoe","Dream",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ bill_length_mm   : num [1:344] 39.1 39.5 40.3 NA 36.7 39.3 38.9 39.2 34.1 42 ...
##  $ bill_depth_mm    : num [1:344] 18.7 17.4 18 NA 19.3 20.6 17.8 19.6 18.1 20.2 ...
##  $ flipper_length_mm: int [1:344] 181 186 195 NA 193 190 181 195 193 190 ...
##  $ body_mass_g      : int [1:344] 3750 3800 3250 NA 3450 3650 3625 4675 3475 4250 ...
##  $ sex              : Factor w/ 2 levels "female","male": 2 1 1 NA 1 2 1 2 NA NA ...
##  $ year             : int [1:344] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 ...
```

Running the code to get the scatterplot to show the relationship between flipper length and body mass of three penguin species

```
ggplot(data=penguins)+
  geom_point(mapping=aes(x=flipper_length_mm, y=body_mass_g, color=species))+
  labs(title="Palmer Penguins: Body Mass Vs Flipper Length", subtitle="Sample of Three Penguin Species"
       caption="Data Collected by Dr. Kristen Gorman")+
  annotate("text", x=220, y=3500, label="The Gentoos are the largest",
           color="purple", fontface="bold", size=4.5, angle=25)
```

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



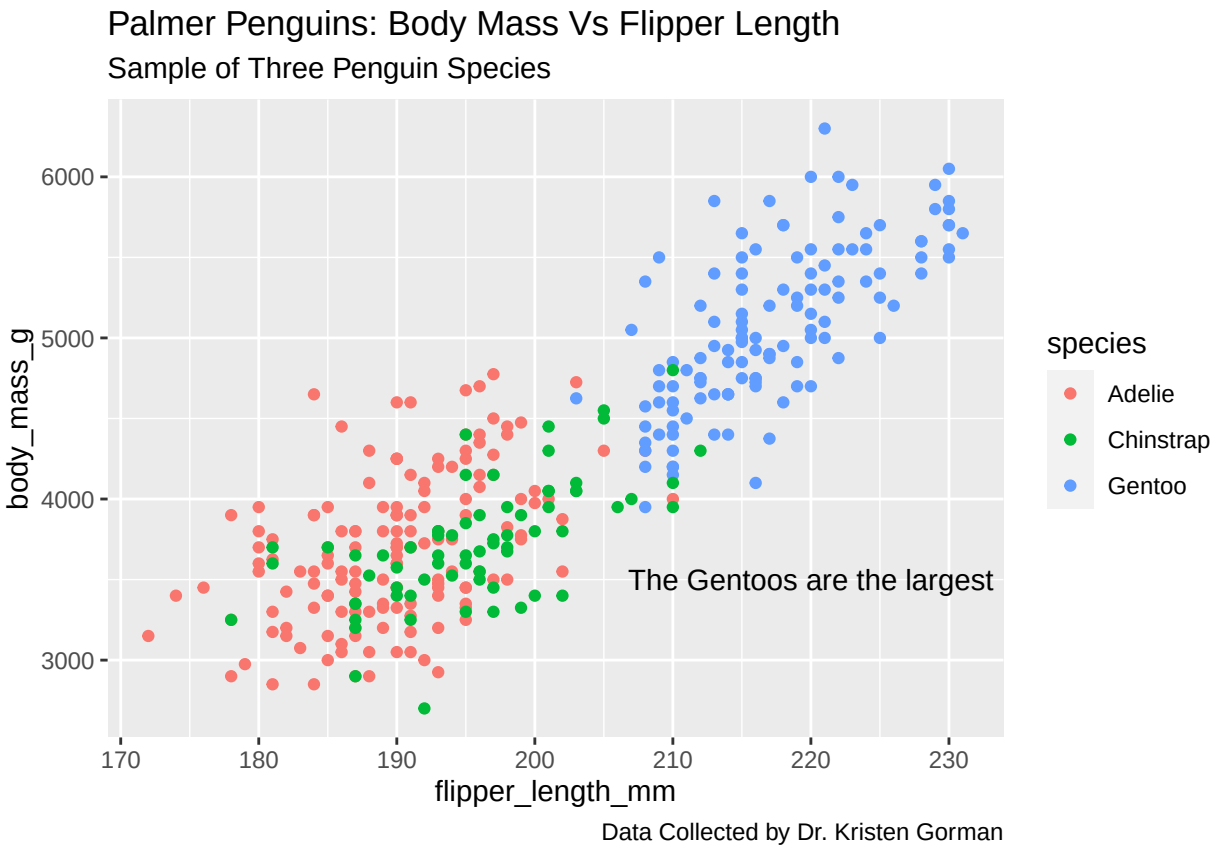
Breaking my code by assigning a portion to a new variable 'p' to reduce the length of my codes

```
p <- ggplot(data=penguins)+
  geom_point(mapping=aes(x=flipper_length_mm, y=body_mass_g, color=species))+
  labs(title="Palmer Penguins: Body Mass Vs Flipper Length", subtitle="Sample of Three Penguin Species"
       caption="Data Collected by Dr. Kristen Gorman")
```

Join the variable 'p' with the code showing some annotations on the plot

```
p + annotate("text", x=220,y=3500, label="The Gentoos are the largest")
```

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



Conclusion: The Gentoos species are the largest among the three species and they have longer flipper length and body mass compared to Adelie and Chunstrap