Penguins' Species

Author: Aimal Khan

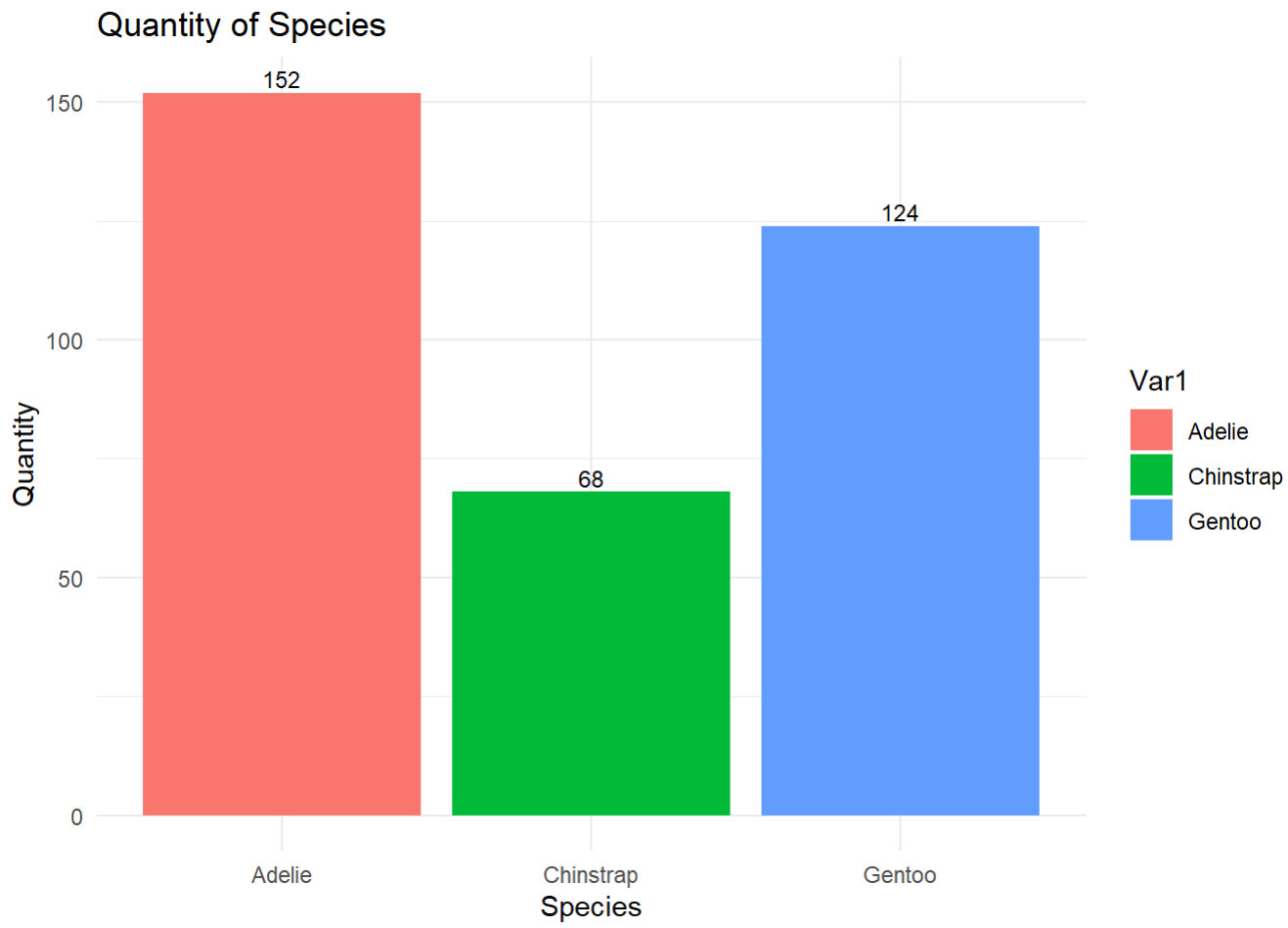
Dated: 2024-02-02

```
library(tidyverse)
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.3     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.4     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all confl
icts to become errors
library(palmerpenguins)
## Warning: package 'palmerpenguins' was built under R version 4.3.2
## Statistical Summary
summary(penguins)
##       species          island    bill_length_mm  bill_depth_mm
##  Adelie   :152   Biscoe   :168   Min.   :32.10   Min.   :13.10
##  Chinstrap: 68   Dream    :124   1st Qu.:39.23   1st Qu.:15.60
##  Gentoo   :124   Torgersen: 52   Median :44.45   Median :17.30
##                                  Mean   :43.92   Mean   :17.15
##                                  3rd Qu.:48.50   3rd Qu.:18.70
##                                  Max.   :59.60   Max.   :21.50
##                                  NA's   :2       NA's   :2
##  flipper_length_mm  body_mass_g       sex           year
##  Min.   :172.0     Min.   :2700   female:165   Min.   :2007
##  1st Qu.:190.0     1st Qu.:3550   male  :168   1st Qu.:2007
##  Median :197.0     Median :4050   NA's  : 11   Median :2008
##  Mean   :200.9     Mean   :4202                Mean   :2008
##  3rd Qu.:213.0     3rd Qu.:4750                3rd Qu.:2009
##  Max.   :231.0     Max.   :6300                Max.   :2009
##  NA's   :2         NA's   :2
## Calculate the summary statistics for each species
penguin_summary <- penguins %>%
```

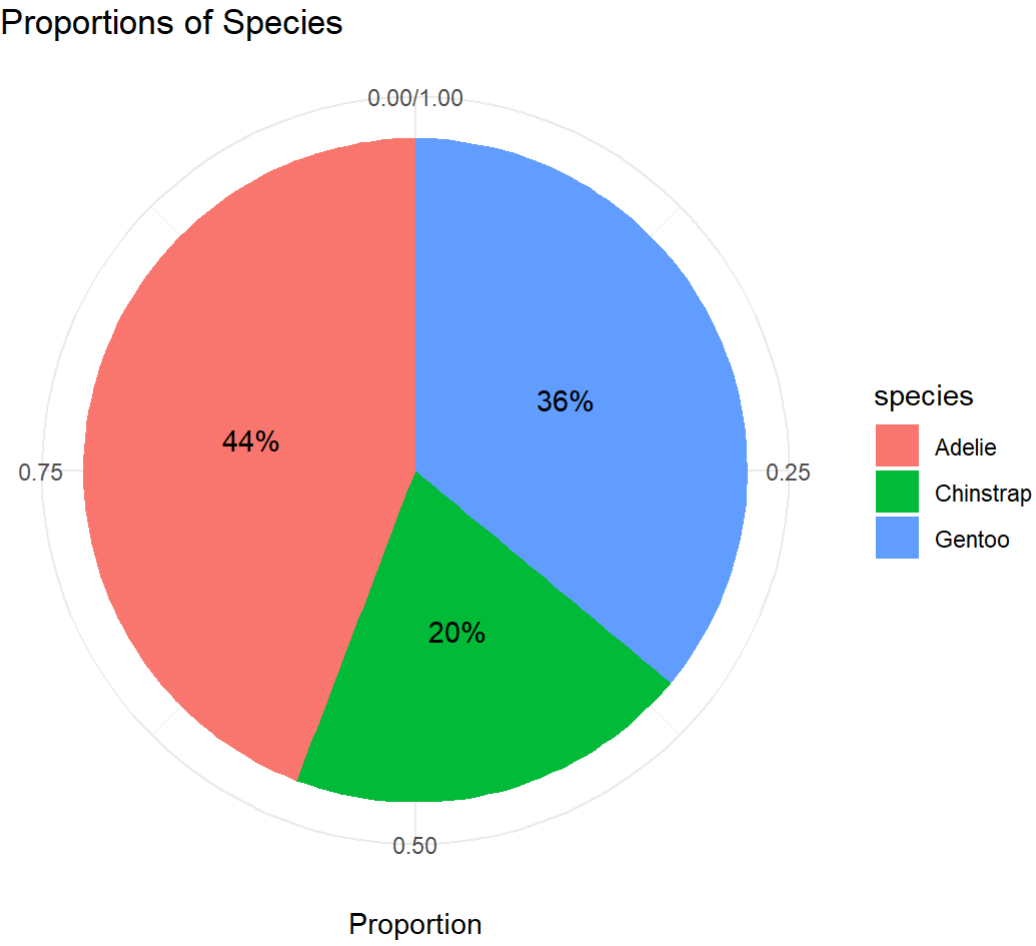
```
  group_by(species) %>%
  summarize(
    n = n(),
    mean_bill_length = mean(bill_length_mm, na.rm = TRUE),
    mean_bill_depth = mean(bill_depth_mm, na.rm = TRUE),
    mean_flipper_length = mean(flipper_length_mm, na.rm = TRUE),
    mean_body_mass = mean(body_mass_g, na.rm = TRUE)
  )
# Print the summary
print(penguin_summary)
## # A tibble: 3 × 6
##   species       n mean_bill_length mean_bill_depth mean_flipper_length
##   <fct>     <int>            <dbl>           <dbl>                <dbl>
## 1 Adelie      152             38.8            18.3                 190.
## 2 Chinstrap    68             48.8            18.4                 196.
## 3 Gentoo      124             47.5            15.0                 217.
## # ℹ 1 more variable: mean_body_mass <dbl>
## Visualizations
# The total number of observations and proportions:
total_obs <- nrow(penguins)
proportions <- penguins %>%
  group_by(species) %>%
  summarize(
    n = n(),
    prop = n / total_obs
  )

# The total number of observations:
total_obs <- nrow(penguins)
species_counts <- table(penguins$species)
```

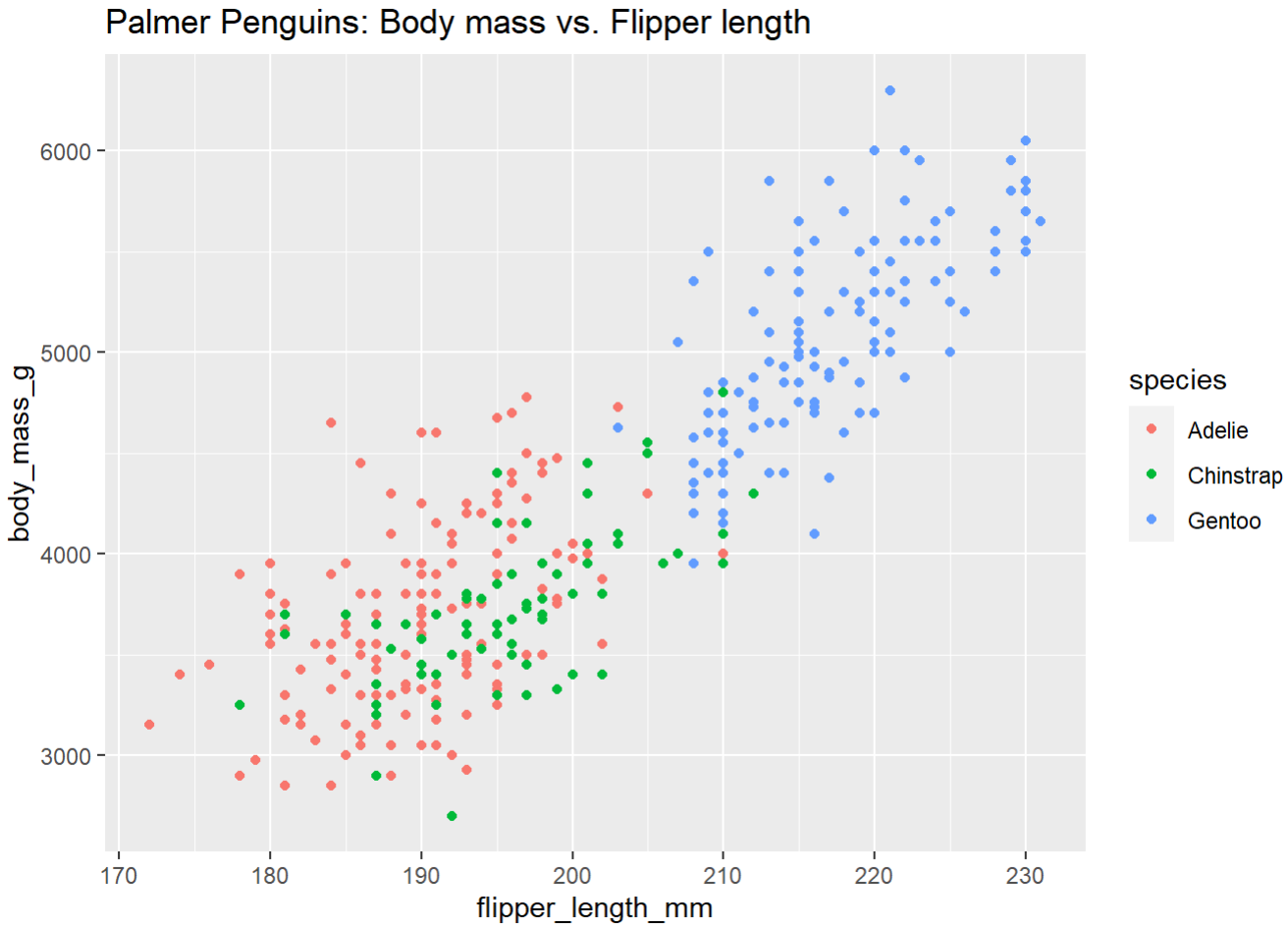
```r
# Quantity of Species (Bar chart)
library(ggplot2)
ggplot(data = as.data.frame(species_counts), aes(x = Var1, y = Freq, fill = Var1))
+
  geom_bar(stat = "identity") +
  labs(title = "Quantity of Species ", x = "Species", y = "Quantity") +
  theme_minimal() +
  geom_text(aes(label = Freq), vjust = -0.3, size = 3)
```



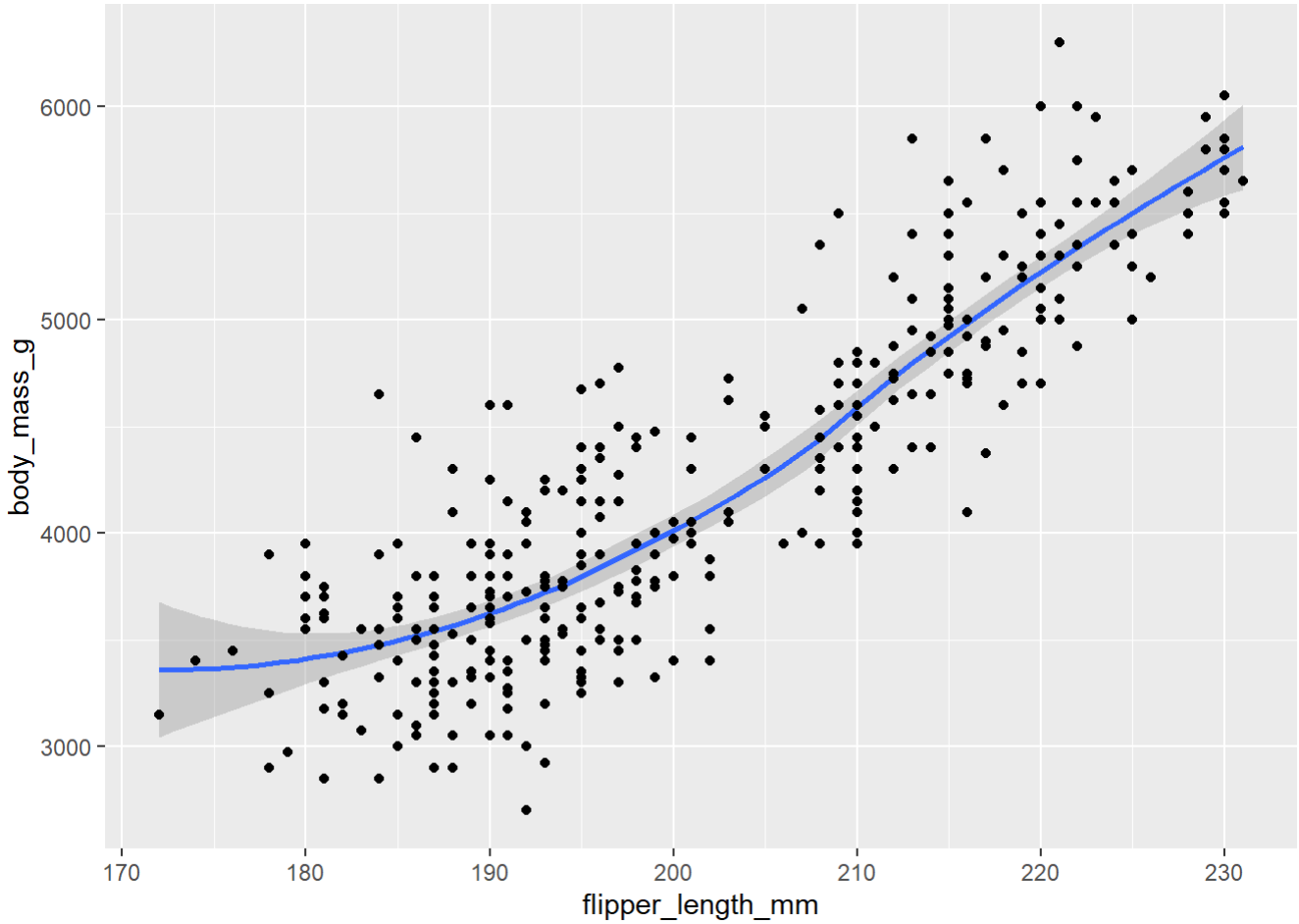
```
# A pie chart with labels
library(ggplot2)
ggplot(proportions, aes(x = "", y = prop, fill = species)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar(theta = "y") +
  labs(title = "Proportions of Species", x = "", y = "Proportion") +
  theme_minimal() +
  geom_text(aes(label = paste0(round(prop * 100), "%")), position = position_stack(
vjust = 0.5))
```



```
ggplot(data=penguins)+geom_point(mapping=aes(x=flipper_length_mm,y=body_mass_g,colo
r=species))+
  labs(title = "Palmer Penguins: Body mass vs. Flipper length")
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```

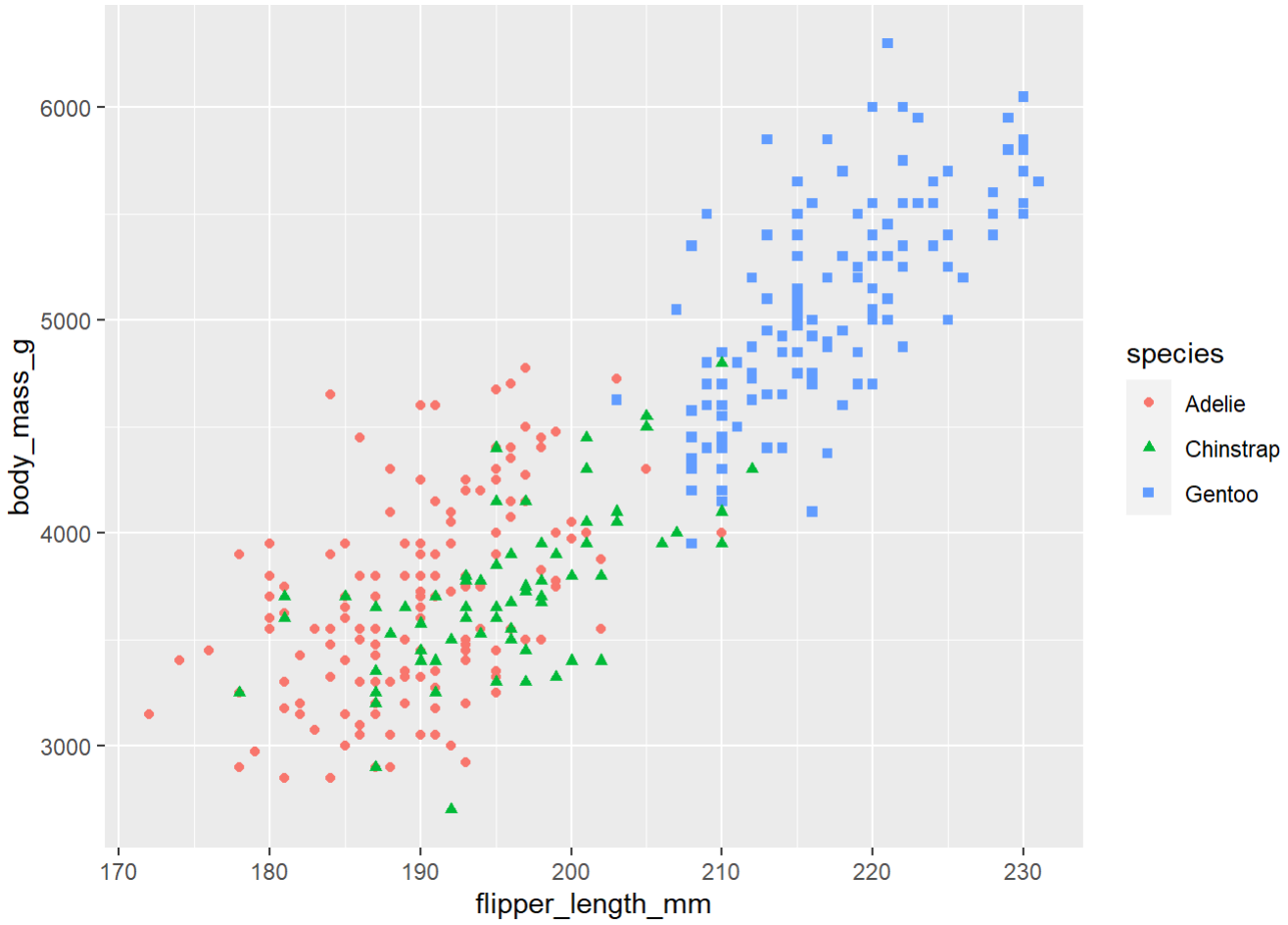


```
ggplot(data=penguins)+geom_smooth(aes(x=flipper_length_mm,y=body_mass_g))+
  geom_point(aes(x=flipper_length_mm,y=body_mass_g))
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
## Removed 2 rows containing missing values (`geom_point()`).
```



```
ggplot(data=penguins)+geom_point(mapping=aes(x=flipper_length_mm,y=body_mass_g,shape=species,color=species))

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



```
ggplot(data=penguins)+geom_point(mapping=aes(x=flipper_length_mm,y=body_mass_g,colo
r=species))+
  labs(title = "Palmer Penguins: Body mass vs. Flipper length",
       subtitle = "Sample of Three Penguin Species", caption = "Data Collected by D
r. Kristen Gorman")
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```