Assignment 3 Solutions

Data Avicenna

2025-06-03

Setup

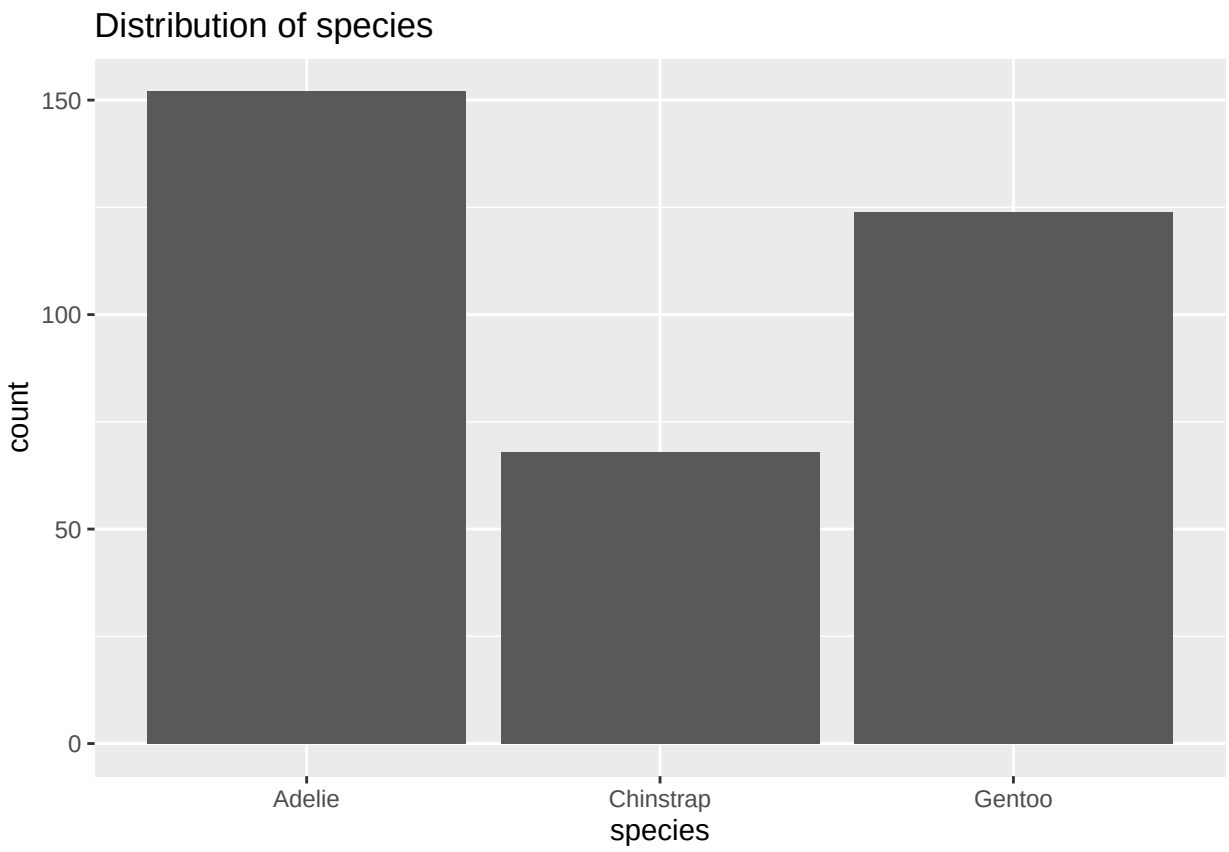
Use palmerpenguins and ggplot2 package:

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

Question 1

Bar graph:

```
ggplot(penguins, aes(species)) +
  geom_bar() +
  labs(
    title = "Distribution of species")
```



```r
# Check number of species (optional)
library(tidyverse)
```

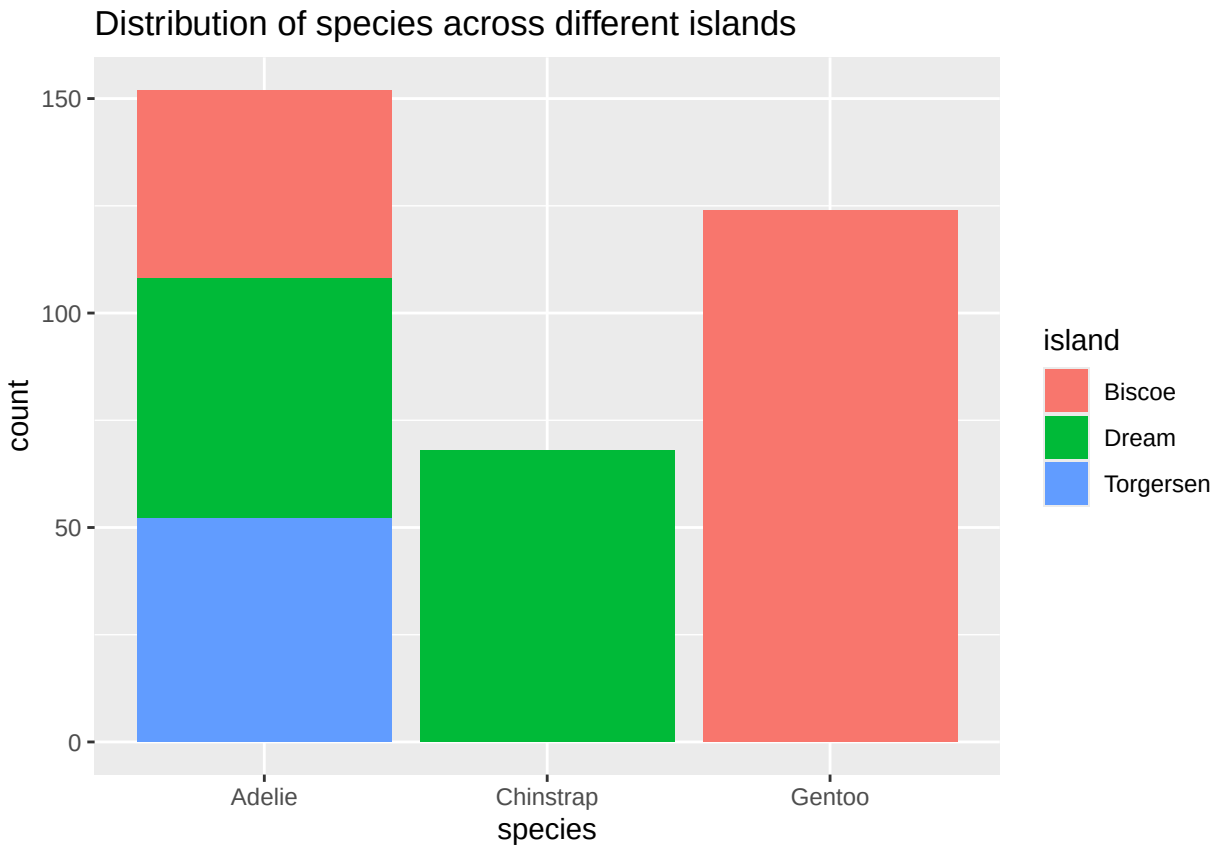
```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v lubridate 1.9.4     v tibble    3.2.1
## v purrr     1.0.4     v tidyr     1.3.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
penguins %>%
    count(species)
```

```
## # A tibble: 3 x 2
##   species       n
##   <fct>     <int>
## 1 Adelie      152
## 2 Chinstrap    68
## 3 Gentoo      124
```

Stacked bar chart:

```r
ggplot(penguins, aes(species, fill = island)) +
  geom_bar(position = "stack") +
   labs(
    title = "Distribution of species across different islands")
```



In the penguins dataset, the highest number of observations of species is Adelie, corresponding to 152 observations. Then, there is Gentoo and Chinstrap in which each has 124 and 68 observations.

Visualizing the distribution of species across different islands, we find that the Adelie species is distributed across all three islands, while the Chinstrap species only reside in Dream island and the Gentoo species only reside in Biscoe island.

Question 2

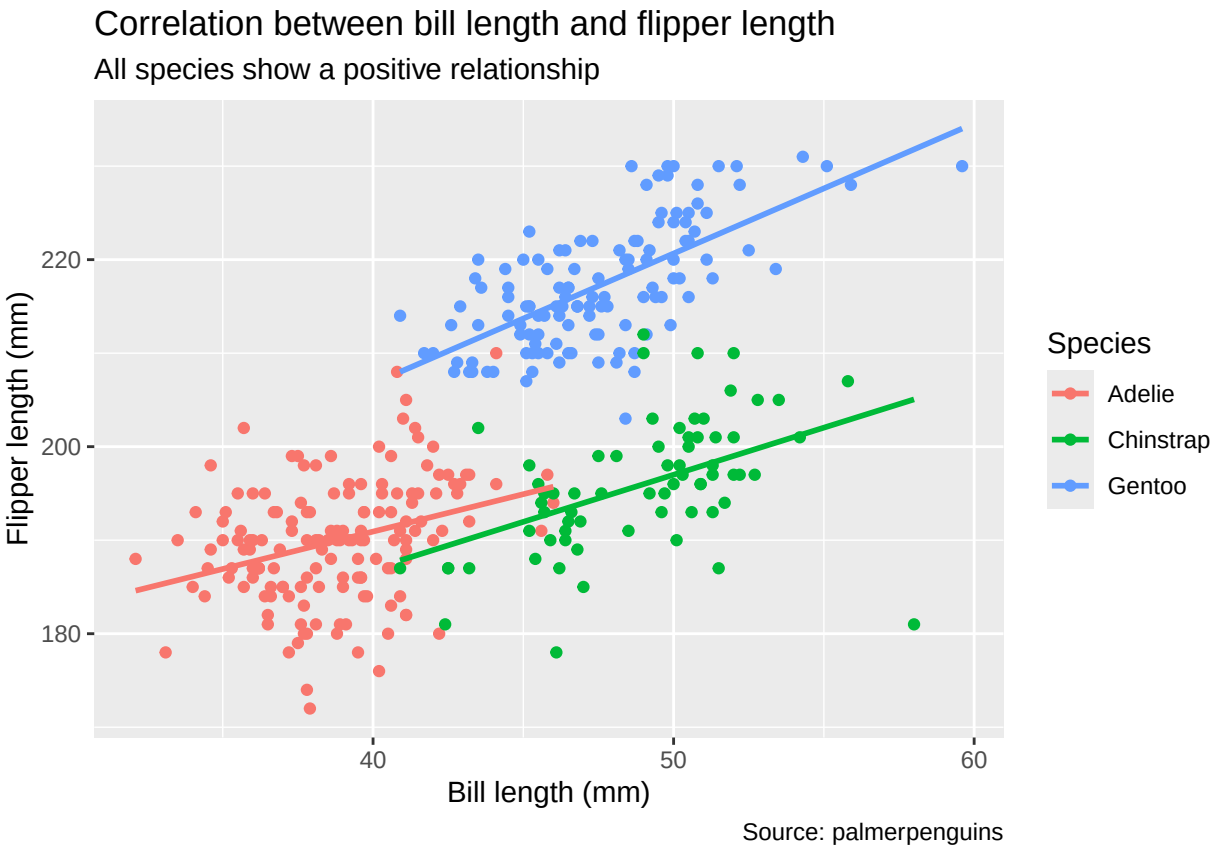
```
ggplot(penguins, aes(x = bill_length_mm,
                     y = flipper_length_mm,
                     color = species)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
labs(
    title = "Correlation between bill length and flipper length",
    subtitle = "All species show a positive relationship",
    y = "Flipper length (mm)",
    x = "Bill length (mm)",
    color = "Species",
    caption = "Source: palmerpenguins"
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 2 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
```

```
## (`geom_point()`).
```



All species of penguins show a positive relationship between bill length and flipper length.