Penguin Story: A Visual Exploration

Ehsna Zangene

Meet the Penguins

The Palmer Penguins dataset contains size measurements, clutch observations, and blood isotope ratios for three penguin species: **Adelie**, **Chinstrap**, and **Gentoo**, observed on three islands in the Palmer Archipelago, Antarctica (2007–2009).

This project explores their story through tidy data and reproducible R code.

```
penguins_raw <- read_csv("https://raw.githubusercontent.com/esnzgn/DataViz-R-Course/main/data/penguins_
```

```
## Rows: 344 Columns: 17
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (10): study_name, species, region, island, stage, individual_id, clutch_...
## dbl  (7): sample_number, culmen_length_mm, culmen_depth_mm, flipper_length_m...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(penguins_raw)
```

```
## # A tibble: 6 x 17
##   study_name sample_number species             region island stage individual_id
##   <chr>              <dbl> <chr>               <chr>  <chr>  <chr> <chr>
## 1 PAL0708                1 Adelie Penguin (Py~ Anvers Torge~ Adul~ N1A1
## 2 PAL0708                2 Adelie Penguin (Py~ Anvers Torge~ Adul~ N1A2
## 3 PAL0708                3 Adelie Penguin (Py~ Anvers Torge~ Adul~ N2A1
## 4 PAL0708                4 Adelie Penguin (Py~ Anvers Torge~ Adul~ N2A2
## 5 PAL0708                5 Adelie Penguin (Py~ Anvers Torge~ Adul~ N3A1
## 6 PAL0708                6 Adelie Penguin (Py~ Anvers Torge~ Adul~ N3A2
## # i 10 more variables: clutch_completion <chr>, date_egg <chr>,
## #   culmen_length_mm <dbl>, culmen_depth_mm <dbl>, flipper_length_mm <dbl>,
## #   body_mass_g <dbl>, sex <chr>, delta_15_n_o_oo <dbl>, delta_13_c_o_oo <dbl>,
## #   comments <chr>
```

Cleaning the Data

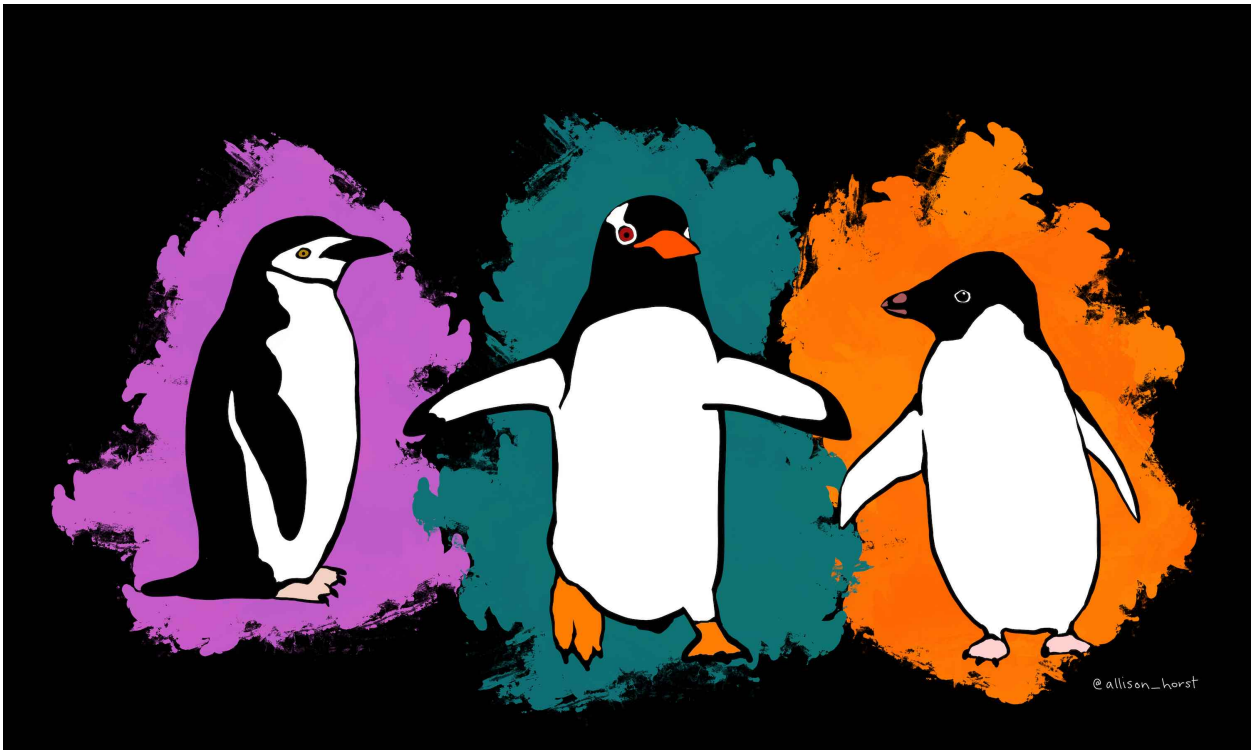


Figure 1: "Penguin Body"

```
penguins <- penguins_raw |>
  filter(!is.na(species)) |>
  mutate(sex = factor(sex), species = factor(species), island = factor(island))

penguins
```

```
## # A tibble: 344 x 17
##    study_name sample_number species            region island stage individual_id
##    <chr>              <dbl> <fct>              <chr>  <fct>  <chr> <chr>
##  1 PAL0708                1 Adelie Penguin (P~ Anvers Torge~ Adul~ N1A1
##  2 PAL0708                2 Adelie Penguin (P~ Anvers Torge~ Adul~ N1A2
##  3 PAL0708                3 Adelie Penguin (P~ Anvers Torge~ Adul~ N2A1
##  4 PAL0708                4 Adelie Penguin (P~ Anvers Torge~ Adul~ N2A2
##  5 PAL0708                5 Adelie Penguin (P~ Anvers Torge~ Adul~ N3A1
##  6 PAL0708                6 Adelie Penguin (P~ Anvers Torge~ Adul~ N3A2
##  7 PAL0708                7 Adelie Penguin (P~ Anvers Torge~ Adul~ N4A1
##  8 PAL0708                8 Adelie Penguin (P~ Anvers Torge~ Adul~ N4A2
##  9 PAL0708                9 Adelie Penguin (P~ Anvers Torge~ Adul~ N5A1
## 10 PAL0708               10 Adelie Penguin (P~ Anvers Torge~ Adul~ N5A2
## # i 334 more rows
## # i 10 more variables: clutch_completion <chr>, date_egg <chr>,
## #   culmen_length_mm <dbl>, culmen_depth_mm <dbl>, flipper_length_mm <dbl>,
## #   body_mass_g <dbl>, sex <fct>, delta_15_n_o_oo <dbl>, delta_13_c_o_oo <dbl>,
## #   comments <chr>
```

Figure 2: "Penguins Species"

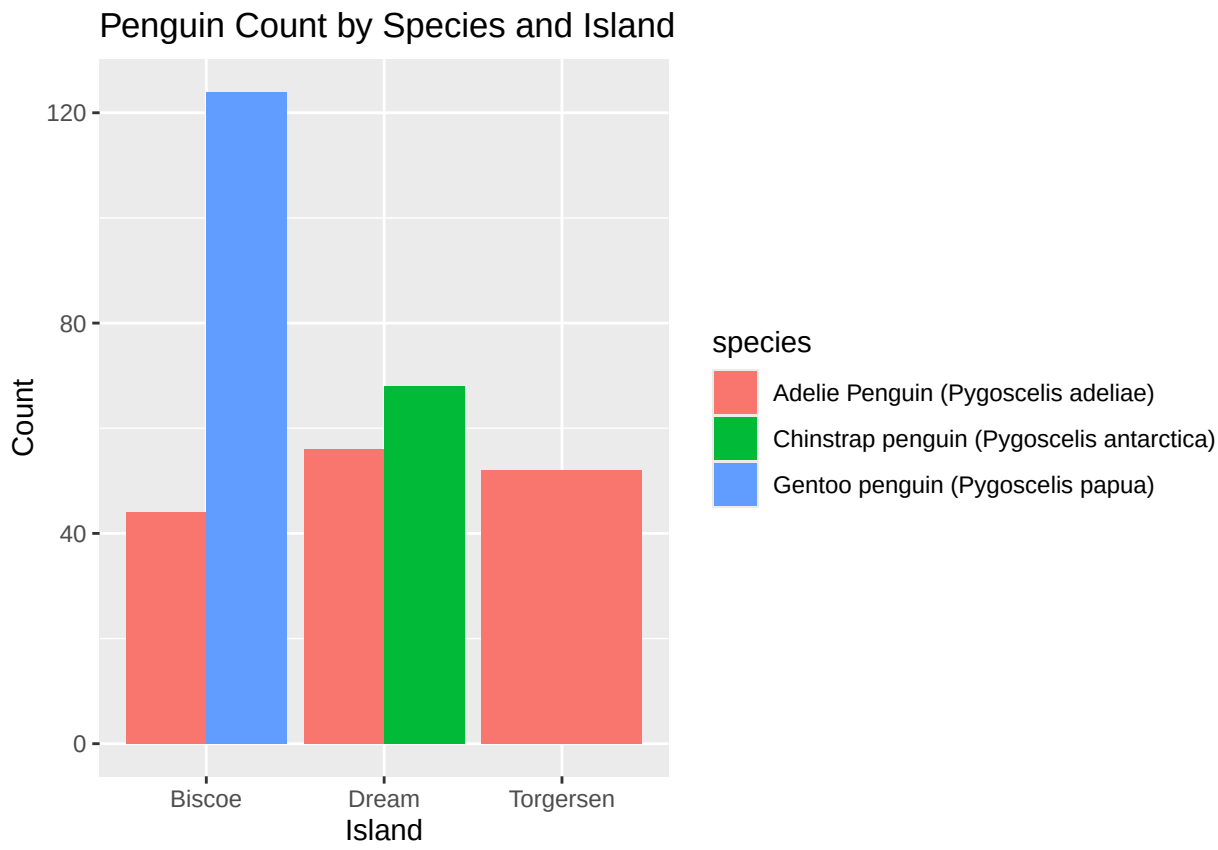
Summary Statistics

```
summary(penguins)
```

```
##   study_name        sample_number
##  Length:344         Min.   :  1.00
##  Class :character   1st Qu.: 29.00
##  Mode  :character   Median : 58.00
##                     Mean   : 63.15
##                     3rd Qu.: 95.25
##                     Max.   :152.00
##
##                                       species       region
##  Adelie Penguin (Pygoscelis adeliae)      :152   Length:344
##  Chinstrap penguin (Pygoscelis antarctica): 68   Class :character
##  Gentoo penguin (Pygoscelis papua)        :124   Mode  :character
##
##
##
##
##        island       stage           individual_id      clutch_completion
##  Biscoe   :168   Length:344         Length:344         Length:344
##  Dream    :124   Class :character   Class :character   Class :character
##  Torgersen: 52   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    date_egg         culmen_length_mm culmen_depth_mm flipper_length_mm
##  Length:344         Min.   :32.10    Min.   :13.10   Min.   :172.0
##  Class :character   1st Qu.:39.23    1st Qu.:15.60   1st Qu.:190.0
##  Mode  :character   Median :44.45    Median :17.30   Median :197.0
##                     Mean   :43.92    Mean   :17.15   Mean   :200.9
##                     3rd Qu.:48.50    3rd Qu.:18.70   3rd Qu.:213.0
##                     Max.   :59.60    Max.   :21.50   Max.   :231.0
##                     NA's   :2        NA's   :2       NA's   :2
##   body_mass_g       sex      delta_15_n_o_oo  delta_13_c_o_oo
##  Min.   :2700   FEMALE:165   Min.   : 7.632   Min.   :-27.02
##  1st Qu.:3550   MALE  :168   1st Qu.: 8.300   1st Qu.:-26.32
##  Median :4050   NA's  : 11   Median : 8.652   Median :-25.83
##  Mean   :4202                Mean   : 8.733   Mean   :-25.69
##  3rd Qu.:4750                3rd Qu.: 9.172   3rd Qu.:-25.06
##  Max.   :6300                Max.   :10.025   Max.   :-23.79
##  NA's   :2                   NA's   :14       NA's   :13
##    comments
##  Length:344
##  Class :character
##  Mode  :character
##
##
##
##
```

Penguin Population by Species and Island

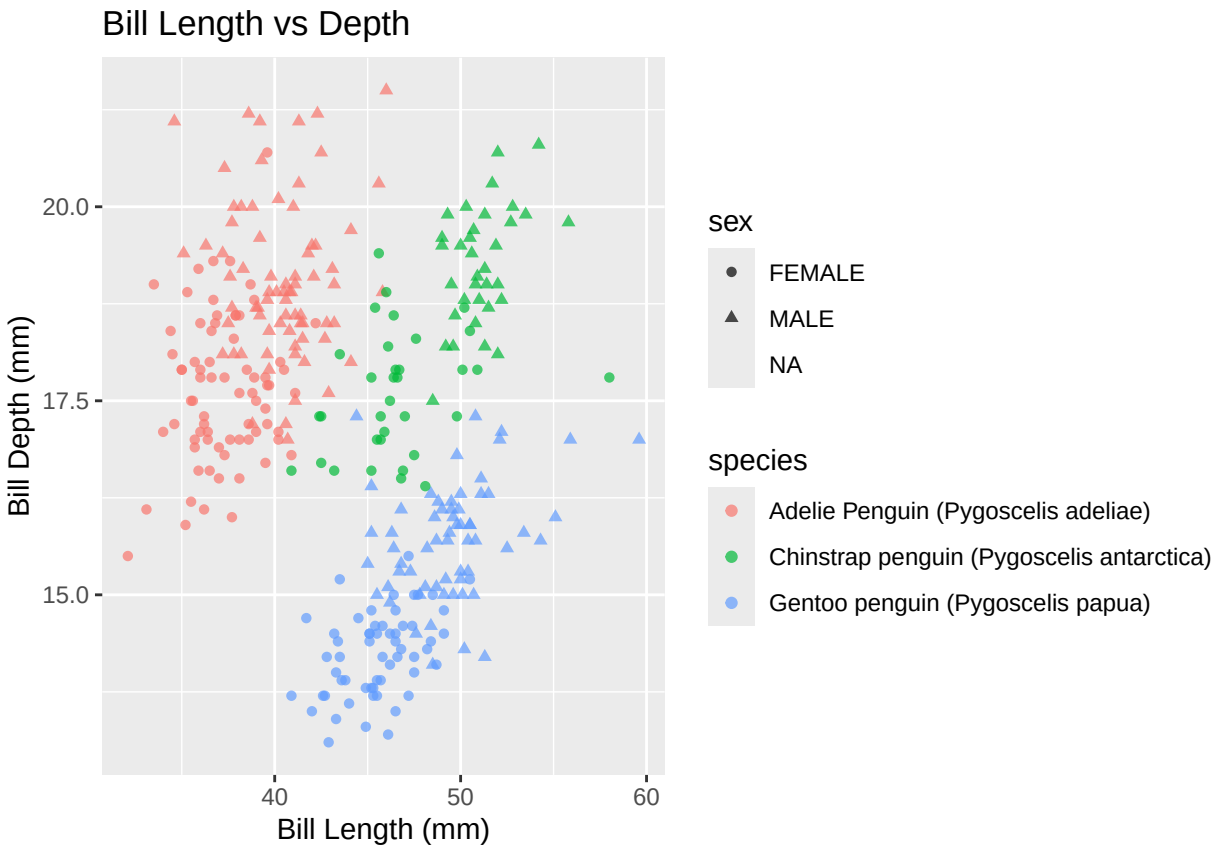
```
penguins |>
  count(species, island) |>
  ggplot(aes(x = island, y = n, fill = species)) +
  geom_col(position = "dodge") +
  labs(title = "Penguin Count by Species and Island", x = "Island", y = "Count")
```



Bill Length vs Depth

```
ggplot(penguins, aes(x = culmen_length_mm, y = culmen_depth_mm, color = species, shape = sex)) +
  geom_point(alpha = 0.7) +
  labs(title = "Bill Length vs Depth", x = "Bill Length (mm)", y = "Bill Depth (mm)")
```

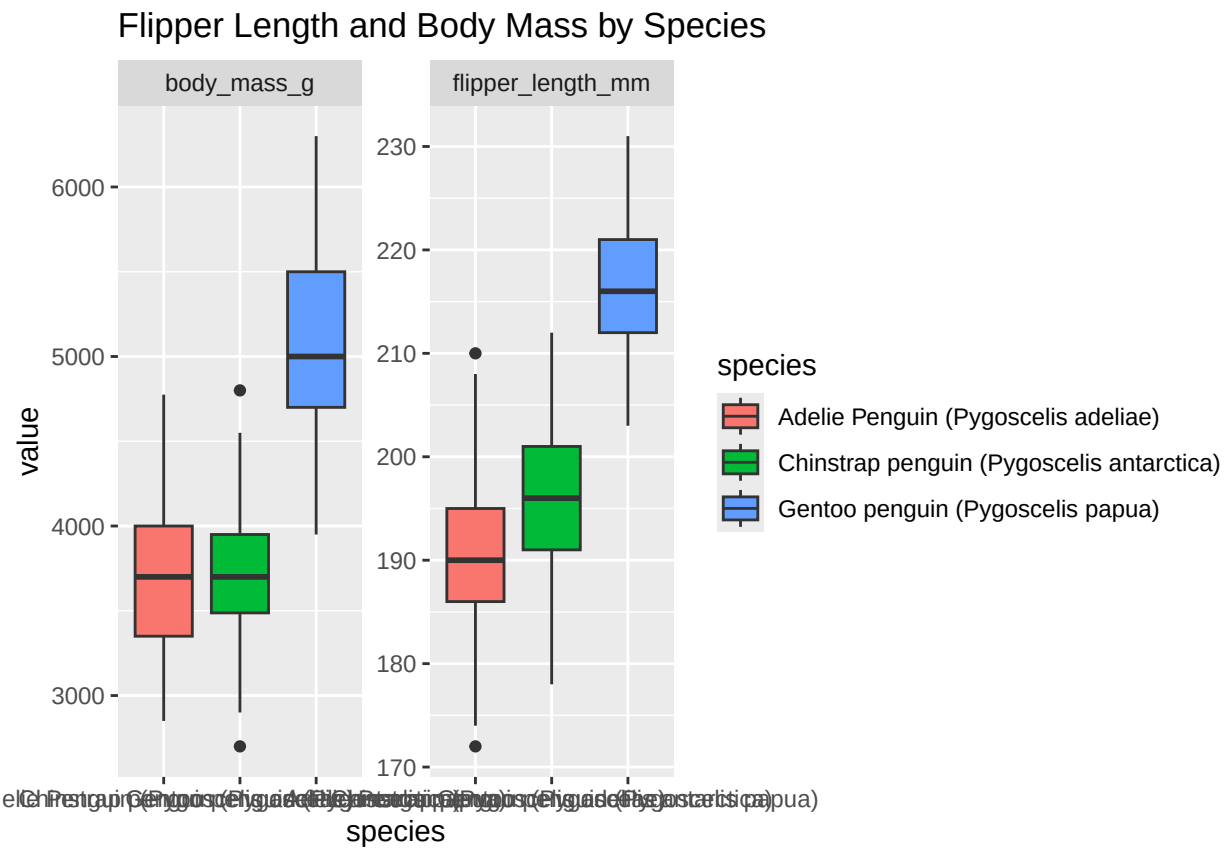
```
## Warning: Removed 11 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



Flipper Length and Body Mass by Species

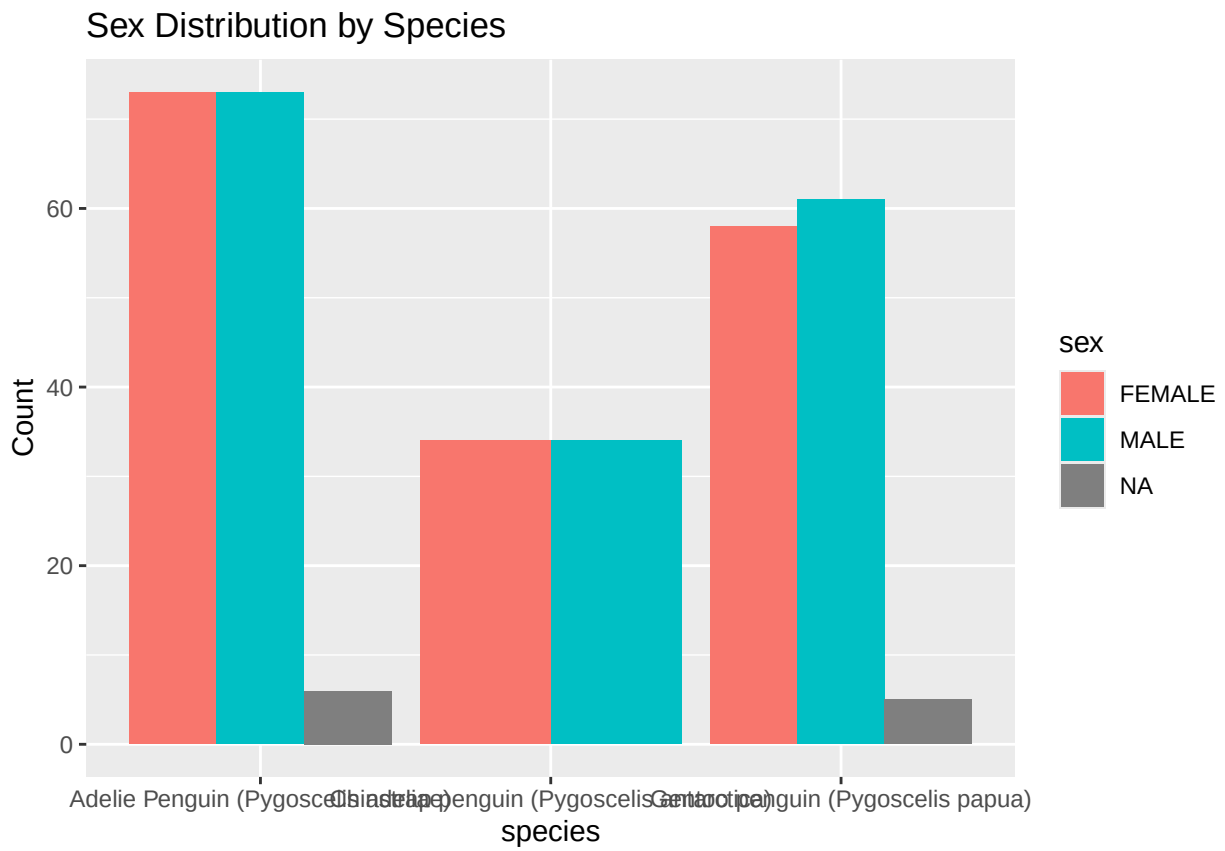
```
penguins |>
  pivot_longer(c(flipper_length_mm, body_mass_g), names_to = "trait", values_to = "value") |>
  ggplot(aes(x = species, y = value, fill = species)) +
  geom_boxplot() +
  facet_wrap(~trait, scales = "free_y") +
  labs(title = "Flipper Length and Body Mass by Species")
```

```
## Warning: Removed 4 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



Sex Distribution

```
penguins |>
  count(sex, species) |>
  ggplot(aes(x = species, y = n, fill = sex)) +
  geom_col(position = "dodge") +
  labs(title = "Sex Distribution by Species", y = "Count")
```



Conclusion

The Palmer Penguins offer a beautiful story of biodiversity, adaptation, and scientific discovery. Using R, we visualize and explore how size, sex, species, and geography shape this Antarctic tale.

Data from Gorman et al. (2014) via palmerpenguins R package.