

# Palmer Penguins

Kenny

2024-09-14

Analyzing a dataset called Palmer Penguins. The palmerpenguins data contains size measurements for three penguin species observed on three islands in the Palmer Archipelago, Antarctica.

```
library(palmerpenguins)
library(skimr)
library(skimr)
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
## ✓ forcats    1.0.0      ✓ stringr    1.5.1
## ✓ ggplot2    3.5.1      ✓ tibble     3.2.1
## ✓ lubridate  1.9.3      ✓ tidyr      1.3.1
## ✓ 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 conflicts to become errors
```

```
library(dplyr)
library(ggplot2)
library(skimr)
library(tidyverse)
```

Get a summary of the data (1)

```
skim_without_charts(penguins)
```

## Data summary

Name	penguins
Number of rows	344
Number of columns	8
Column type frequency:	
factor	3
numeric	5

---

Group variables

None

### Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
species	0	1.00	FALSE	3	Ade: 152, Gen: 124, Chi: 68
island	0	1.00	FALSE	3	Bis: 168, Dre: 124, Tor: 52
sex	11	0.97	FALSE	2	mal: 168, fem: 165

### Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
bill_length_mm	2	0.99	43.92	5.46	32.1	39.23	44.45	48.5	59.6
bill_depth_mm	2	0.99	17.15	1.97	13.1	15.60	17.30	18.7	21.5
flipper_length_mm	2	0.99	200.92	14.06	172.0	190.00	197.00	213.0	231.0
body_mass_g	2	0.99	4201.75	801.95	2700.0	3550.00	4050.00	4750.0	6300.0
year	0	1.00	2008.03	0.82	2007.0	2007.00	2008.00	2009.0	2009.0

Get a summary of the data (2)

```
colnames(penguins)
```

```
## [1] "species"      "island"        "bill_length_mm"
## [4] "bill_depth_mm" "flipper_length_mm" "body_mass_g"
## [7] "sex"          "year"
```

Get a summary of the data (3)

```
glimpse(penguins)
```

```
## Rows: 344
## Columns: 8
## $ species      <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adel...
## $ island       <fct> Torgersen, Torgersen, Torgersen, Torgersen, Torgerse...
## $ bill_length_mm <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1, ...
## $ bill_depth_mm <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1, ...
## $ flipper_length_mm <int> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 186...
## $ body_mass_g   <int> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475, ...
## $ sex          <fct> male, female, female, NA, female, male, female, male...
## $ year          <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007...
```

Counting the number of species shows us there are 3 in total as shown from the skim function above

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

species <fct>	n <int>
Adelie	152
Chinstrap	68
Gentoo	124
3 rows	

Leads dive into some calculation

Mean bill length my sex. Male penguins have longer and deeper bills, longer flippers, and weight more on average.

```
penguins %>%
  group_by(sex) %>%
  summarise(mean(bill_length_mm), mean(bill_depth_mm), mean(flipper_length_mm),mean(body_mass_
g))
```

sex <fct>	mean(bill_length_mm) <dbl>	mean(bill_depth_mm) <dbl>	mean(flipper_length_mm) <dbl>	mean(body_
female	42.09697	16.42545	197.3636	
male	45.85476	17.89107	204.5060	
NA	NA	NA	NA	
3 rows				

The Island of Biscoe has the penguins wit the most mass

```
penguins %>%
  drop_na() %>%
  group_by(island) %>%
  summarise(mean(body_mass_g))
```

island <fct>	mean(body_mass_g) <dbl>
Biscoe	4719.172
Dream	3718.902
Torgersen	3708.511
3 rows	

Lets test the correlation between bill length and bill depth. Appears to be weak and negative.

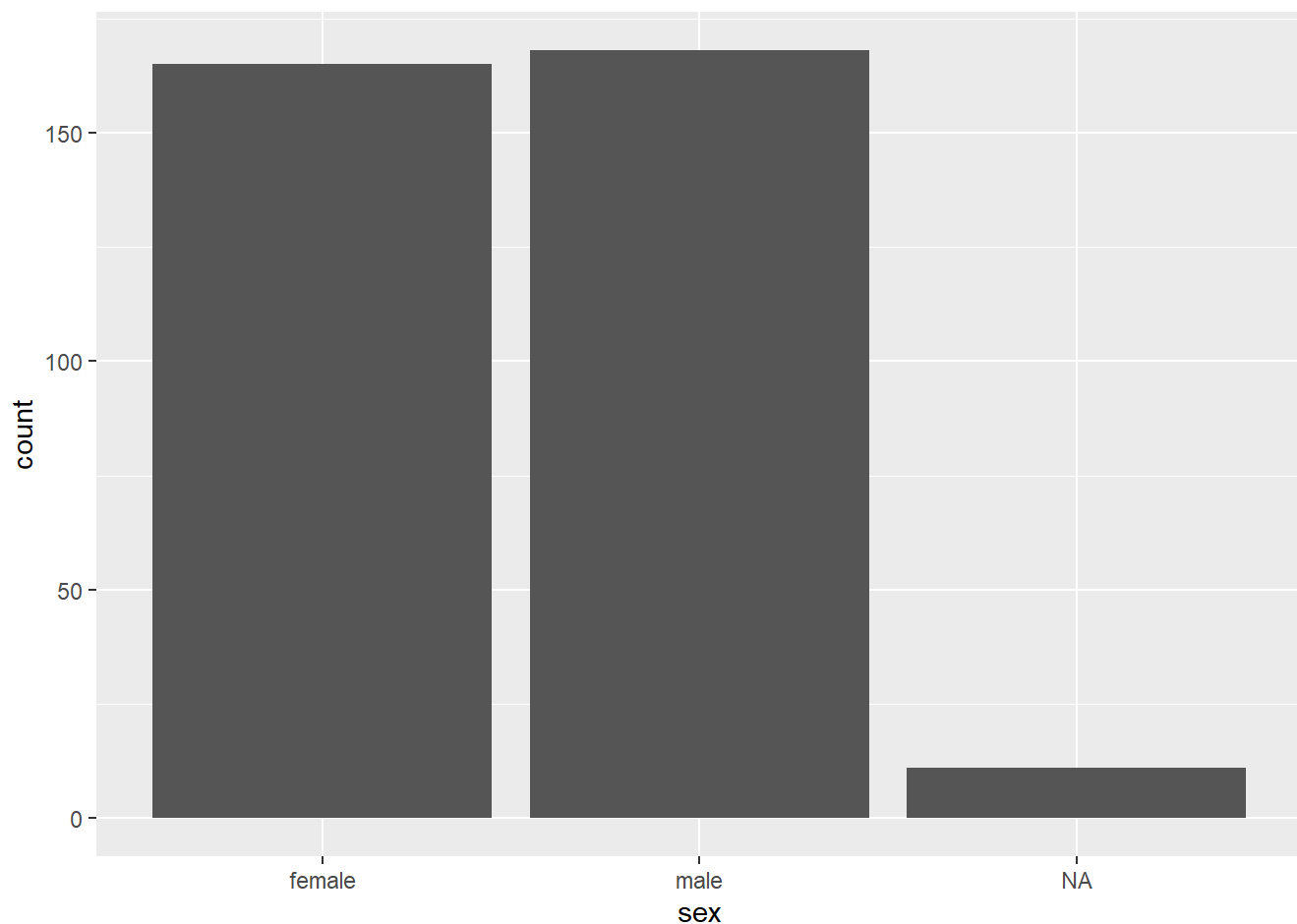
```
bl <- as.numeric(penguins$bill_length_mm)
bd <- as.numeric(penguins$bill_depth_mm)

cor.test(bl,bd)
```

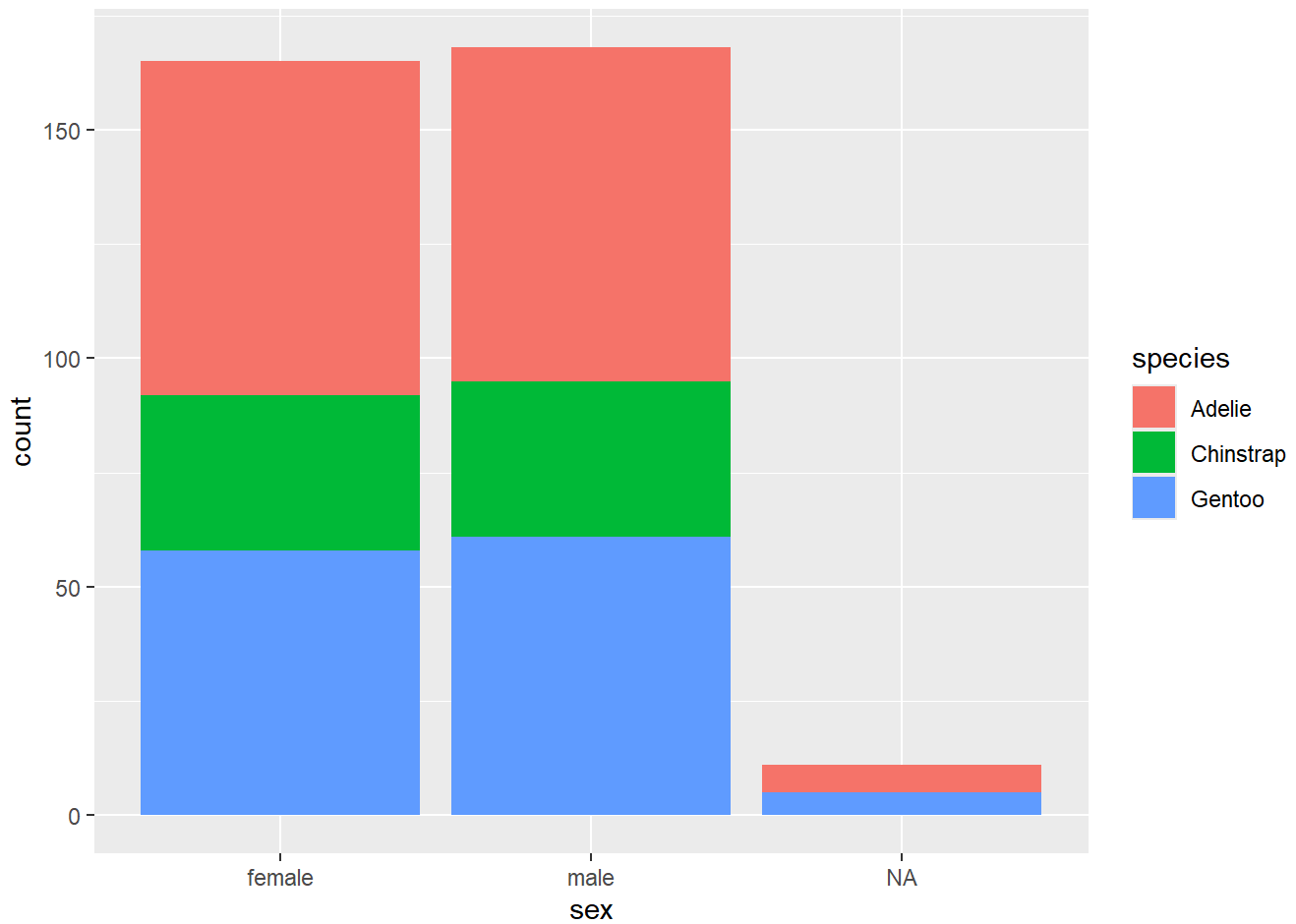
```
##
## Pearson's product-moment correlation
##
## data:  bl and bd
## t = -4.4591, df = 340, p-value = 1.12e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.3328072 -0.1323004
## sample estimates:
##           cor
## -0.2350529
```

Lets plot some data

```
ggplot(data = penguins) +
  geom_bar(mapping = aes(x = sex))
```

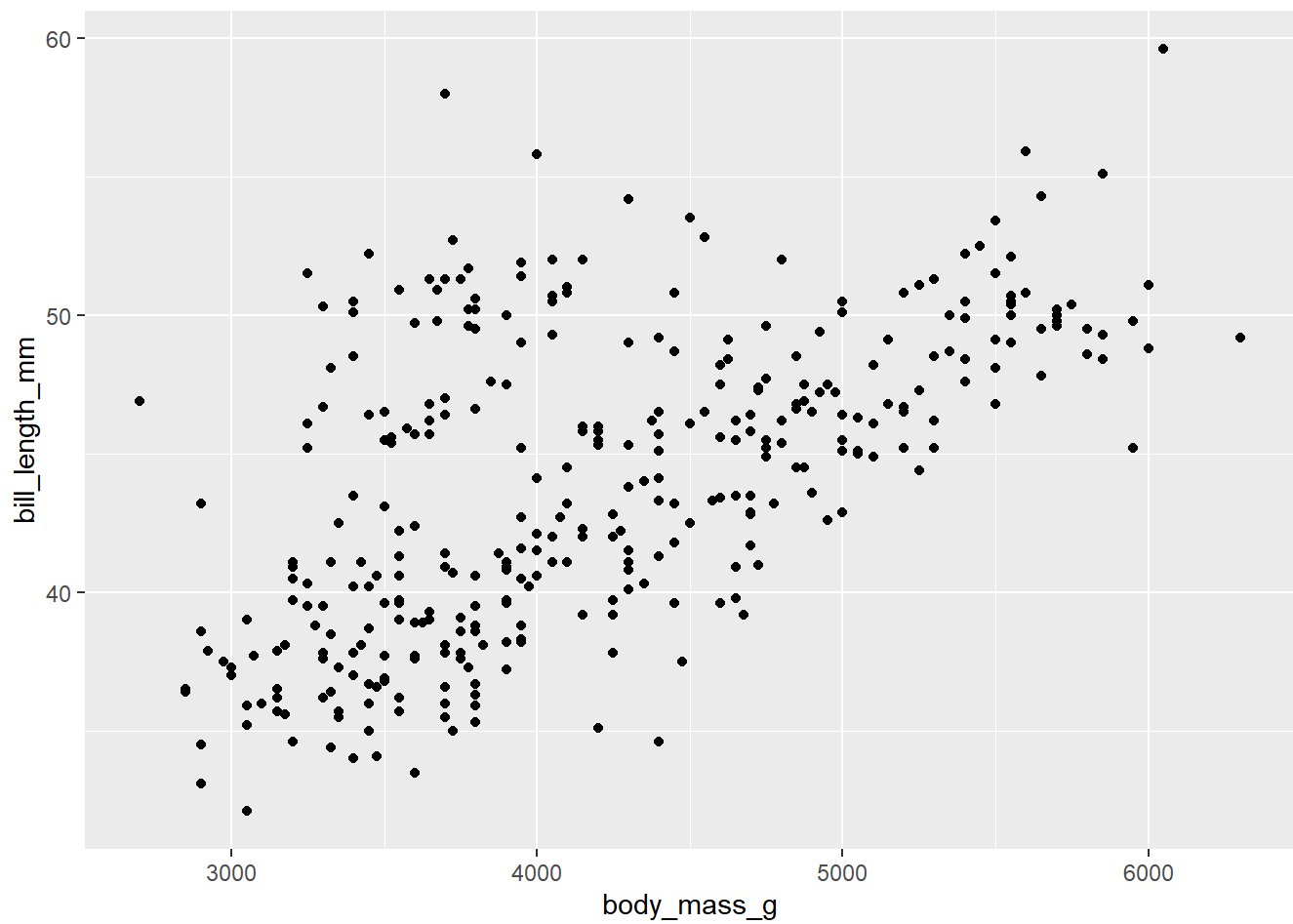


```
ggplot(data = penguins) +  
  geom_bar(mapping = aes(x = sex, fill = species))
```



```
ggplot(data = penguins) +  
  geom_point(mapping = aes(x = body_mass_g, y = bill_length_mm ))
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range  
## (`geom_point()`).
```



```
ggplot(data = penguins) +  
  geom_point(mapping = aes(x = body_mass_g, y = bill_length_mm, color = island ))+  
  facet_wrap(~island) +  
  labs(title = "Penguin Mass and Bill Length Relationship", subtitle = "By Island")
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range  
## (`geom_point()`).
```

# Penguin Mass and Bill Length Relationship

By Island



```
ggplot(data = penguins) +  
  geom_point(mapping = aes(x = body_mass_g, y = bill_length_mm, color = island ))+  
  facet_wrap(~island) +  
  labs(title = "Penguin Mass and Bill Length Relationship", subtitle = "By Island")
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range  
## (`geom_point()`).
```

# Penguin Mass and Bill Length Relationship

By Island



```
ggplot(data = penguins) +  
  geom_point(mapping = aes(x = body_mass_g, y = bill_length_mm, color = island ))+  
  facet_wrap(~island) +  
  labs(title = "Penguin Mass and Bill Length Relationship", subtitle = "By Island", x = "body mass in grams", y = "bill length in mm")
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range  
## (`geom_point()`).
```



## Penguin Mass and Bill Length Relationship

By Island



Thanks for analyzing with me.