

# Estudo de Caso do Curso Google Data Analytics

Jonas Alexandrino

2024-01-11

## Cenário

Você é um analista de dados júnior que trabalha para um consultor de inteligência de negócios. Considerando que está desempenhando sua função há seis meses, seu chefe sente que você está pronto para ter mais responsabilidades. Ele solicitou que você lidere um projeto para um novo cliente – o trabalho envolverá vários aspectos, desde a definição da tarefa de negócios até a apresentação de suas recomendações baseadas em dados. O cliente é o instituto de pesquisa Palmer Station LTER e à US LTER Network, que possui uma estação de pesquisa dos Estados Unidos localizada na Antártica. O objetivo do projeto Palmer LTER é estudar o ecossistema marinho pelágico a oeste da Península Antártica, incluindo habitats de gelo marinho, oceanografia regional e locais de nidificação terrestre de predadores de aves marinhas.



Ask

## Perguntas de Negócio

Em primeiro lugar, o analista de dados se reuniu com às partes interessadas para definir como o projeto seria um resultado de êxito.

Os Pesquisadores do Palmer Station solitam que você analise os dados coletados de 2007 à 2009, que contém medições de tamanho para três espécies de pinguim (Adelie, Chinstrap e Gentoo) que vivem no arquipélago de Palmer, na Antártica. Eles desejam uma comparação por espécie, da massa corporal, tamanho das nadadeiras e sexo.

- Qual a relação por espécie entre a massa corporal e o comprimento da nadadeira?

- Qual a maior e menor espécie de pinguim?
- Qual a relação por sexo entre a massa corporal e o comprimento da nadadeira?



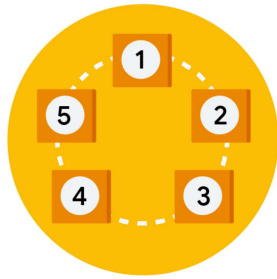
**Prepare**

## Preparar os Dados

Tudo começou com uma preparação sólida. O Analista de dados criou um cronograma de sete dias e como queria transmitir o progresso às partes interessadas. O analista identificou quais dados precisava para atingir o resultado bem-sucedido que identificaram na etapa anterior. Neste caso os dados foram coletados e disponibilizados pela Dra. Kristen Gorman e pela Estação Palmer, Antártica LTER, membro da Rede de Pesquisa Ecológica. O link de acesso: <https://allisonhorst.github.io/palmerpenguins/> (<https://allisonhorst.github.io/palmerpenguins/>). Os dados também estão disponíveis no Pacote R palmerpenguins, que contém dois conjuntos de dados. Para este estudo foi utilizado somente o 'penguins' que é uma versão simplificada dos dados brutos.

Para responder as perguntas de negócios, foram selecionadas as variáveis:

- species: um fator que denota a espécie de pinguim (Adélie, Chinstrap ou Gentoo)
- body\_mass\_g: um inteiro que denota o peso do corpo de um pinguim (gramas)
- flipper\_length\_mm: um inteiro que denota o comprimento de uma nadadeira de pinguim (milímetros)
- sex: um fator que denota o sexo de um pinguim (macho, fêmea) com base em dados moleculares



## Process

# Processar os Dados

Carregando os Pacotes R

```
library("tidyverse")
library("palmerpenguins")
library("skimr")
library("janitor")
library("here")
library("dplyr")
```

Resumo do conjunto de dados

- visualização dos nomes das colunas
- tipos de dados contidos no conjunto de dados

```
head(penguins)
```

species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
<fct>	<fct>	<dbl>	<dbl>	<int>	<int>	<fct>	<int>
Adelie	Torgersen	39.1	18.7	181	3750	male	2007
Adelie	Torgersen	39.5	17.4	186	3800	female	2007

<b>species</b> <fct>	<b>island</b> <fct>	<b>bill_length_mm</b> <dbl>	<b>bill_depth_mm</b> <dbl>	<b>flipper_length_mm</b> <int>	<b>body_mass_g</b> <int>	<b>sex</b> <fct>	<b>year</b> <int>
Adelie	Torgersen	40.3	18.0	195	3250	female	2007
Adelie	Torgersen	NA	NA	NA	NA	NA	2007
Adelie	Torgersen	36.7	19.3	193	3450	female	2007
Adelie	Torgersen	39.3	20.6	190	3650	male	2007
6 rows							

resumo estatístico de cada coluna do conjunto de dados.

- Existem 3 espécies de pinguins: Adelie, Chinstrap e Gentoo
- As medidas do bico variam de 32,10 mm a 59,60 mm
- As medidas de profundidade do bico variam de 13,10 a 21,50 mm
- O comprimento da nadadeira varia de 172,0 mm a 231,0 mm.
- O peso corporal varia de 2700 g a 6300 g.
- Existem 165 fêmeas, 168 machos e 11 valores ausentes para sexo

```
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
```

Garantindo a consistência dos nomes das colunas

- A função `clean_names` no pacote `janitor` vai se certificar automaticamente que os nomes das colunas são únicos e consistentes. Isso garante que só há caracteres, números, e underscores nos nomes

```
clean_names(penguins)
```

<b>species</b> <fct>	<b>island</b> <fct>	<b>bill_length_mm</b> <dbl>	<b>bill_depth_mm</b> <dbl>	<b>flipper_length_mm</b> <int>	<b>body_mass_g</b> <int>	<b>sex</b> <fct>	<b>year</b> <int>
Adelie	Torgersen	39.1	18.7	181	3750	male	2007
Adelie	Torgersen	39.5	17.4	186	3800	female	2007
Adelie	Torgersen	40.3	18.0	195	3250	female	2007
Adelie	Torgersen	NA	NA	NA	NA	NA	2007
Adelie	Torgersen	36.7	19.3	193	3450	female	2007
Adelie	Torgersen	39.3	20.6	190	3650	male	2007

species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year						
<fct>	<fct>	<dbl>	<dbl>	<int>	<int>	<fct>	<int>						
Adelie	Torgersen	38.9	17.8	181	3625	female	2007						
Adelie	Torgersen	39.2	19.6	195	4675	male	2007						
Adelie	Torgersen	34.1	18.1	193	3475	NA	2007						
Adelie	Torgersen	42.0	20.2	190	4250	NA	2007						
1-10 of 344 rows				Previous	1	2	3	4	5	6	...	35	Next

Garantindo que não existe valores duplicados

distinct(penguins)

species <fct>	island <fct>	bill_length_mm <dbl>	bill_depth_mm <dbl>	flipper_length_mm <int>	body_mass_g <int>	sex <fct>	year <int>
Adelie	Torgersen	39.1	18.7	181	3750	male	2007
Adelie	Torgersen	39.5	17.4	186	3800	female	2007
Adelie	Torgersen	40.3	18.0	195	3250	female	2007
Adelie	Torgersen	NA	NA	NA	NA	NA	2007
Adelie	Torgersen	36.7	19.3	193	3450	female	2007
Adelie	Torgersen	39.3	20.6	190	3650	male	2007
Adelie	Torgersen	38.9	17.8	181	3625	female	2007
Adelie	Torgersen	39.2	19.6	195	4675	male	2007
Adelie	Torgersen	34.1	18.1	193	3475	NA	2007
Adelie	Torgersen	42.0	20.2	190	4250	NA	2007

1-10 of 344 rows

Previous123456...35Next



Analyze

## Analisar

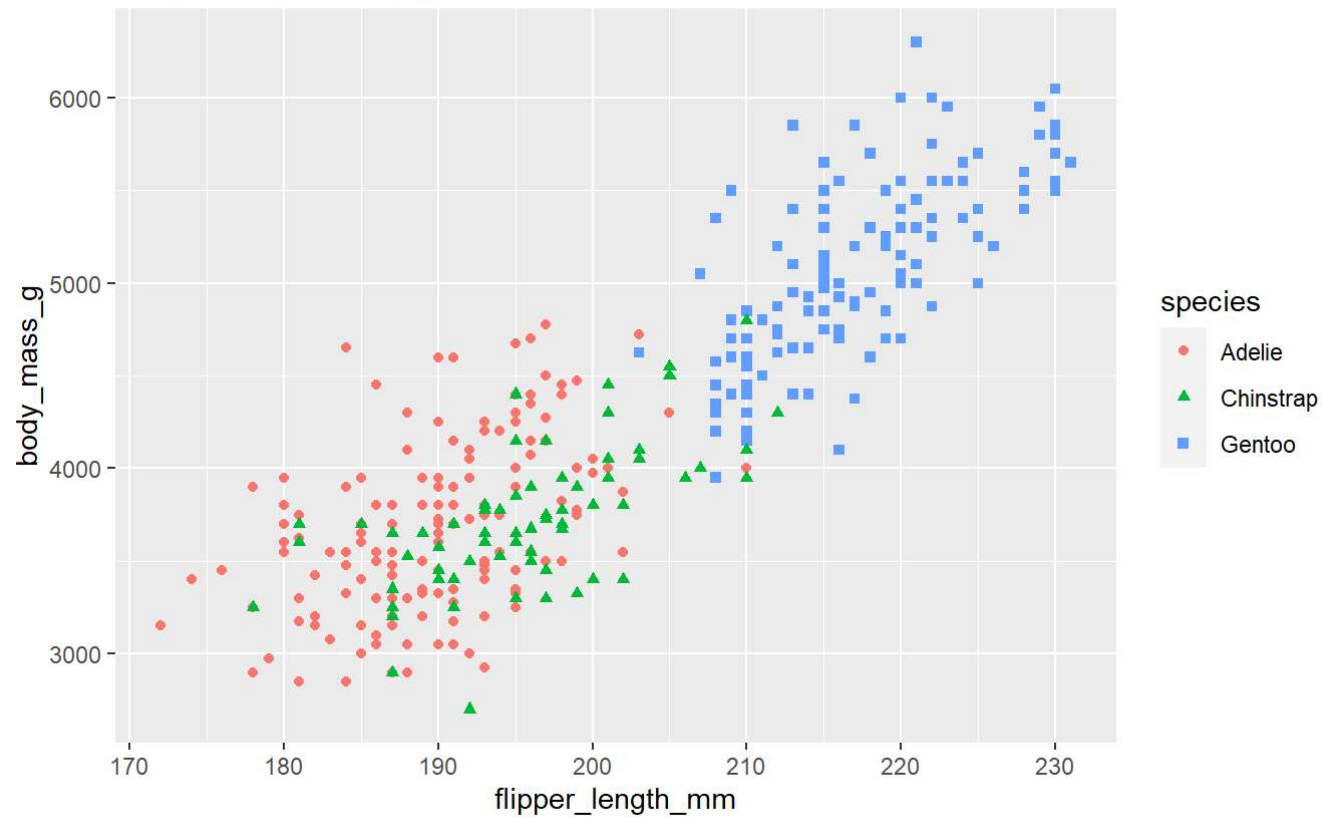
Relação entre o `body_mass_g` (massa corporal) e `flipper_length_mm` (Comprimento da nadadeira)

- O gráfico mostra uma relação positiva entre as duas variáveis. Em outras palavras, quanto maior o pinguim, mais longa é a nadadeira.
- Os Gentoo são os maiores, entre as três espécies.
- Adelie são os menores, entre as três espécies.
- A legenda mostra um círculo para a espécie Adelie, um triângulo para Chinstraps e um quadrado para Gentoo.

```
ggplot(data = penguins) +geom_point(mapping = aes(x = flipper_length_mm, y = body_mass_g,shape = species, color = species))+  
  labs(title="Palmer Penguins: Body Mass vs. Flipper Length", subtitle = "Sample of Three Penguins Species",  
        caption ="Data collected by Dr.Krinten Gorman" )
```

## Palmer Penguins: Body Mass vs. Flipper Length

Sample of Three Penguins Species

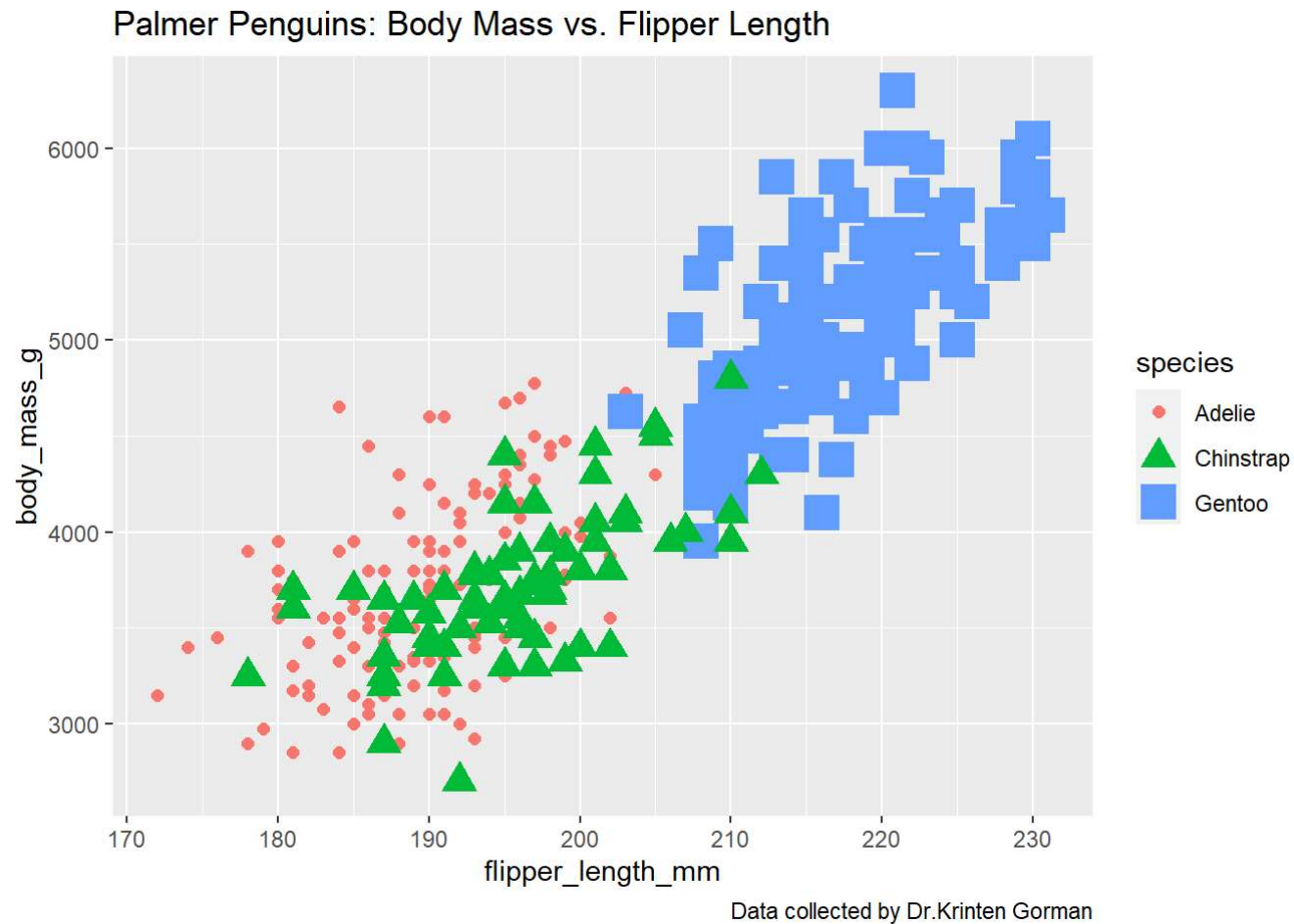


Data collected by Dr.Krinten Gorman

- Outra forma de visualização, destacando o tamanho das formas por espécie

```
ggplot(data = penguins) +geom_point(mapping = aes(x = flipper_length_mm, y = body_mass_g,shape = species, color = species, size = species))+  
  labs(title="Palmer Penguins: Body Mass vs. Flipper Length",  
        caption ="Data collected by Dr.Krinten Gorman" )
```



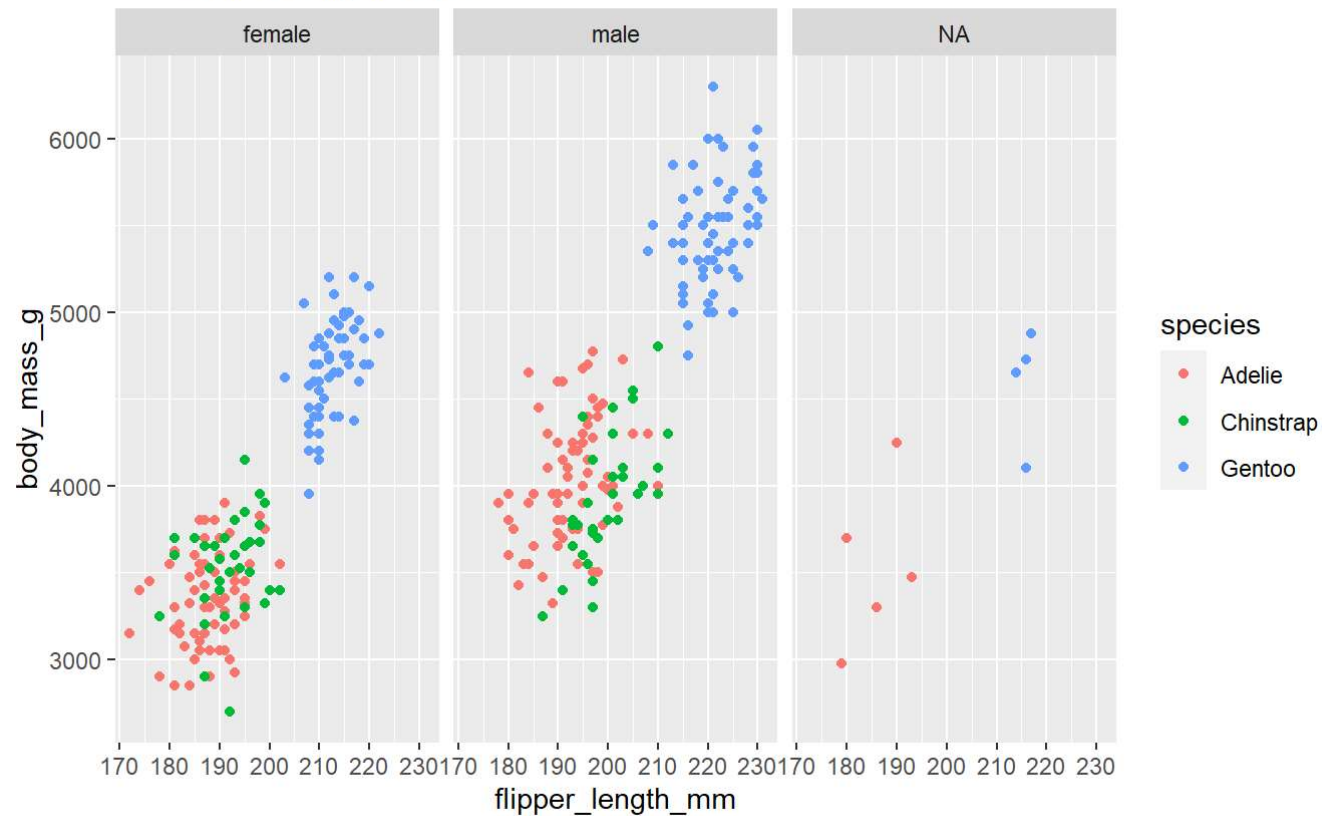


- Os pinguins femininos têm uma tendência de terem menor massa corporal e comprimento de nadadeira em comparação aos masculinos.

```
ggplot(data = penguins) +  
  geom_point(mapping = aes(x = flipper_length_mm, y = body_mass_g,color=species))+  
  facet_grid(~sex)+ labs(title="Palmer Penguins: Body Mass vs. Flipper Length",subtitle = "sex of Three Penguins Species",  
    caption = "Data collected by Dr.Krinten Gorman" )
```

## Palmer Penguins: Body Mass vs. Flipper Length

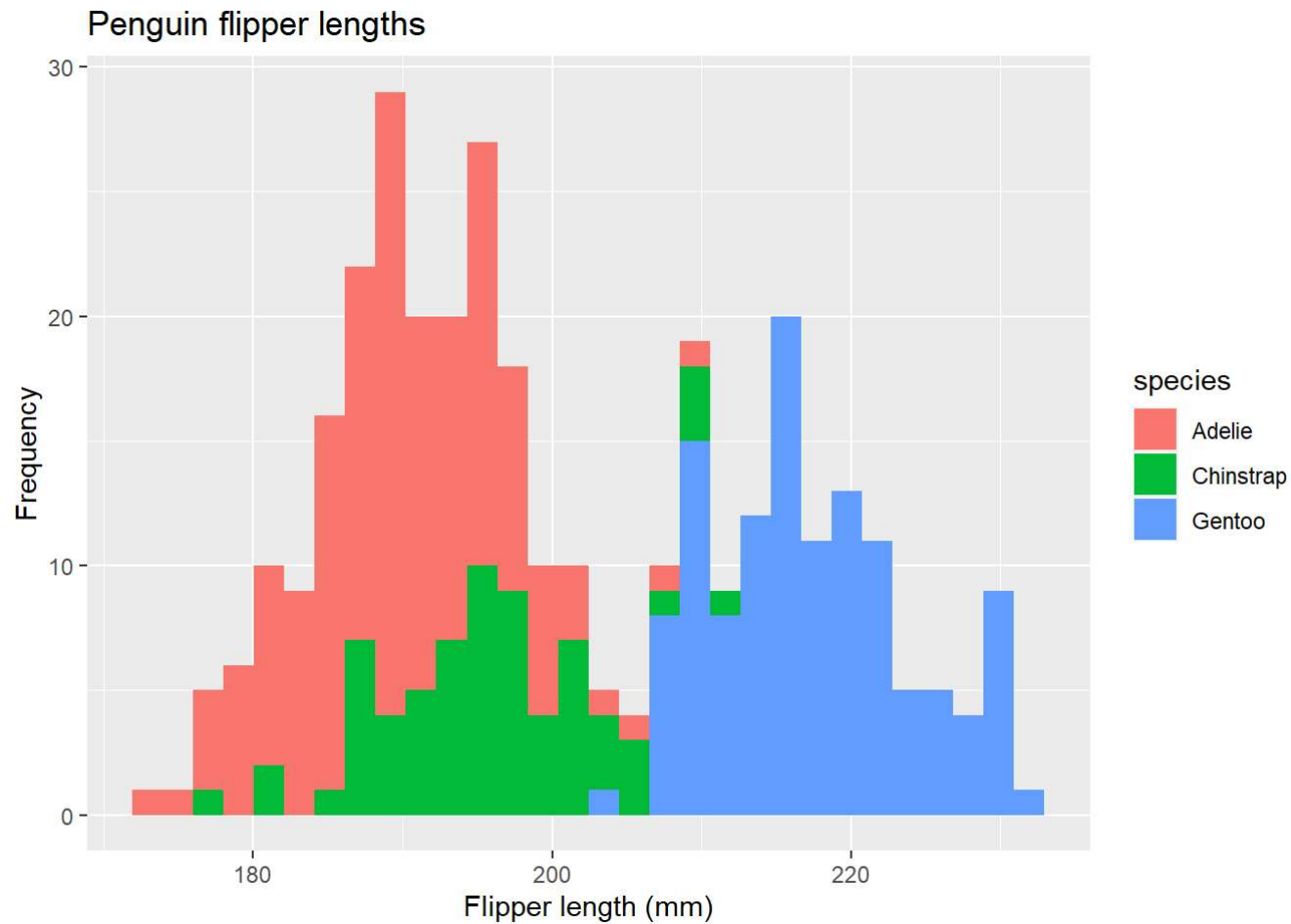
sex of Three Penguins Species



Data collected by Dr.Krinten Gorman

- O Gráfico mostra os comprimentos das asas de três espécies diferentes de pinguins (Adelie, Chinstrap e Gentoo).
- A frequência dos comprimentos das asas é representada no eixo Y, enquanto o comprimento em milímetros está no eixo X.
- Os pinguins Adelie têm comprimentos de asa principalmente entre 180 mm e pouco mais de 200 mm.
- Os pinguins Chinstrap têm uma distribuição semelhante Adelie, mas com menos variação nos comprimentos das asas
- Os pinguins Gentoo têm os maiores comprimentos de asa, variando principalmente entre pouco mais de 200 mm a cerca de 220 mm.

```
ggplot(data = penguins) +  
  geom_histogram(mapping = aes(x = flipper_length_mm, fill = species))+  
  labs(title = "Penguin flipper lengths",  
        x = "Flipper length (mm)",  
        y = "Frequency")
```



- O Gráfico mostra a distribuição da massa corporal de três espécies de pinguins (Adelie, Chinstrap e Gentoo).
- A espécie Adelie tem uma massa corporal que varia principalmente entre 2500g e 4500g. A
- Os pinguins Chinstrap têm uma distribuição semelhante, mas com menos variação nos comprimentos das asas.
- A espécie Gentoo tem uma massa corporal mais elevada, variando principalmente entre 4500g e 6500g.
- O eixo Y representa a frequência dos dados coletados para cada intervalo específico da massa corporal.
- O eixo X representa a massa corporal em gramas.

```
ggplot(data = penguins) +  
  geom_histogram(mapping = aes(x = body_mass_g, fill = species))+  
  labs(title = "Penguin body_mass_g",  
        x = "body mass (g)",  
        y = "Frequency")
```





Share

## Compartilhar

Os resultados foram compartilhados de duas formas às partes interessadas. A primeira foi em uma reunião, utilizando powerpoint, contendo as visualizações das descobertas e principais insights da análise. A segunda forma foi compartilhando um relatório escrito, dando oportunidade de acompanharem as seis etapas do processo de análise de dados: perguntar, preparar, processar, analisar, compartilhar e agir.



Act

## Agir

Conclusão das análises

- Existe uma relação positiva entre as variáveis `bill_length_mm` (comprimento do bico) e `body_mass_g` (massa do corpo). Em outras palavras, quanto maior o pinguim, mais longa é a nadadeira. \*Os pinguins femininos têm uma tendência de terem menor massa corporal e comprimento de nadadeira em comparação aos masculinos.

- Os Gentoos são os maiores pinguins, entre as três espécies.
- Os Adelie são os menores pinguins, entre as três espécies.