

Visualizations of Penguin_dataset

Michael Mbajwa

11/12/2021

```
# Load useful datasets
library(palmerpenguins)
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5      v purrr 0.3.4
## v tibble 3.1.3       v dplyr 1.0.7
## v tidyr 1.1.3        v stringr 1.4.0
## v readr 2.0.1        v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library(ggplot2)
library(hrbrthemes)

## NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use these themes.

##       Please use hrbrthemes::import_roboto_condensed() to install Roboto Condensed and

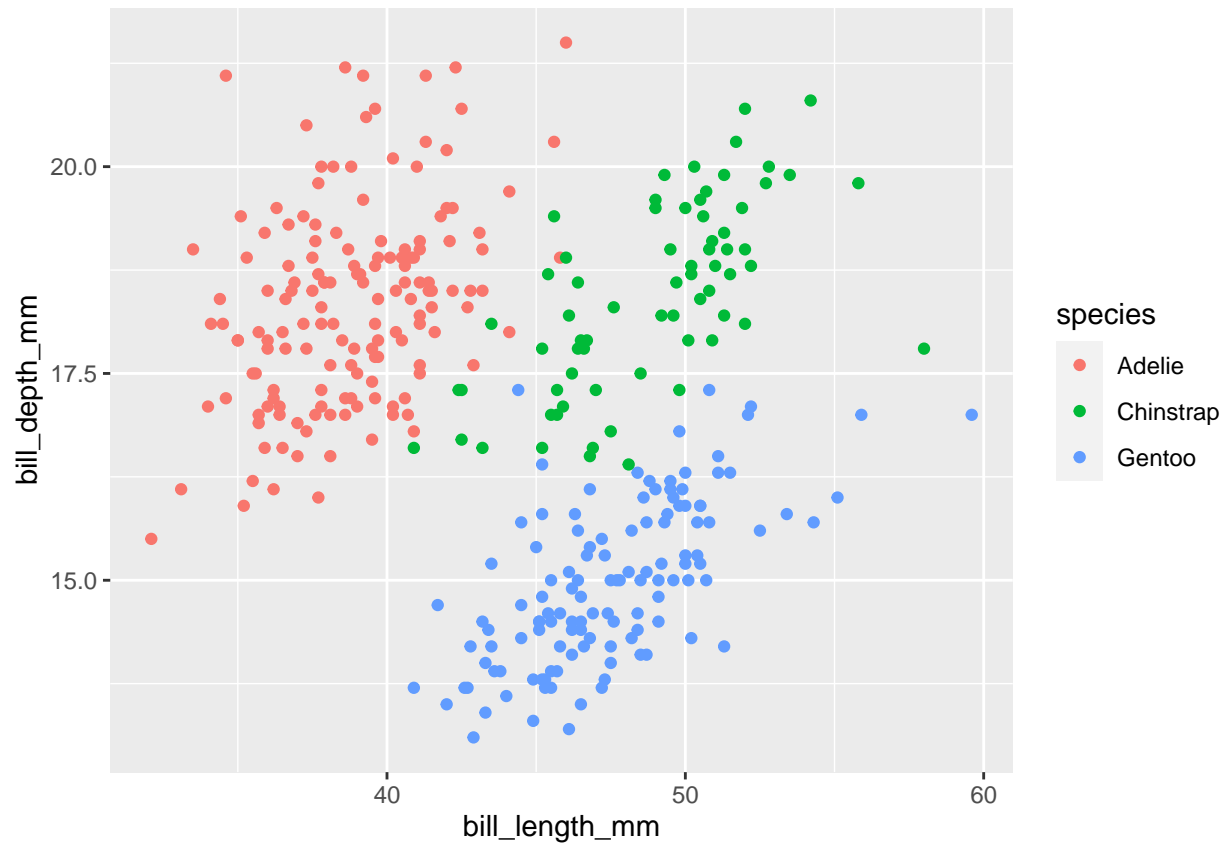
##       if Arial Narrow is not on your system, please see https://bit.ly/arialnarrow

penguins

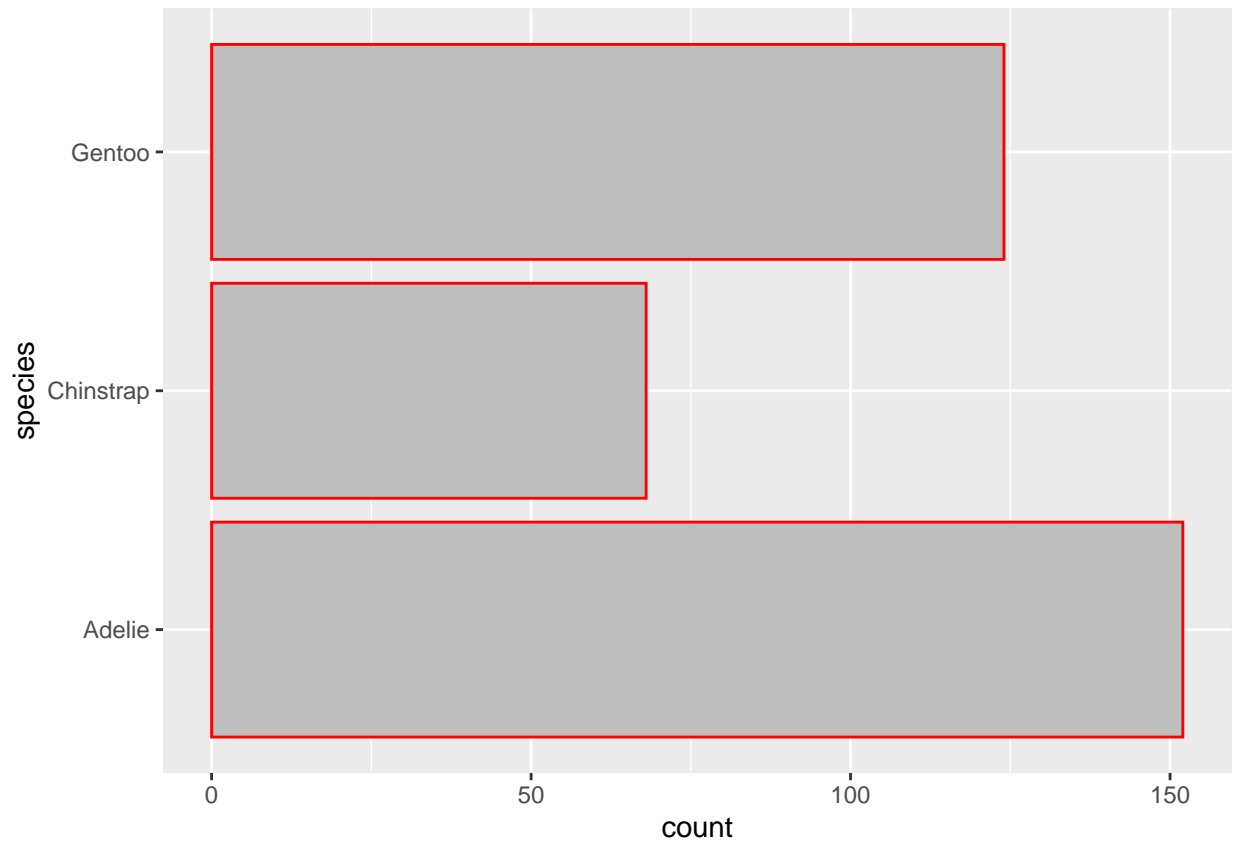
## # A tibble: 344 x 8
##   species island bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
##   <fct>   <fct>         <dbl>         <dbl>         <int>         <int>
## 1 Adelie Torgersen      39.1           18.7           181          3750
## 2 Adelie Torgersen      39.5           17.4           186          3800
## 3 Adelie Torgersen      40.3            18           195          3250
## 4 Adelie Torgersen      NA            NA            NA            NA
## 5 Adelie Torgersen      36.7           19.3           193          3450
## 6 Adelie Torgersen      39.3           20.6           190          3650
## 7 Adelie Torgersen      38.9           17.8           181          3625
## 8 Adelie Torgersen      39.2           19.6           195          4675
## 9 Adelie Torgersen      34.1           18.1           193          3475
## 10 Adelie Torgersen      42            20.2           190          4250
## # ... with 334 more rows, and 2 more variables: sex <fct>, year <int>
```

```
# see relation between bill length and bill depth
penguins %>%
  ggplot() +
    geom_point(mapping = aes(x = bill_length_mm,
                             y = bill_depth_mm,
                             colour = species))
```

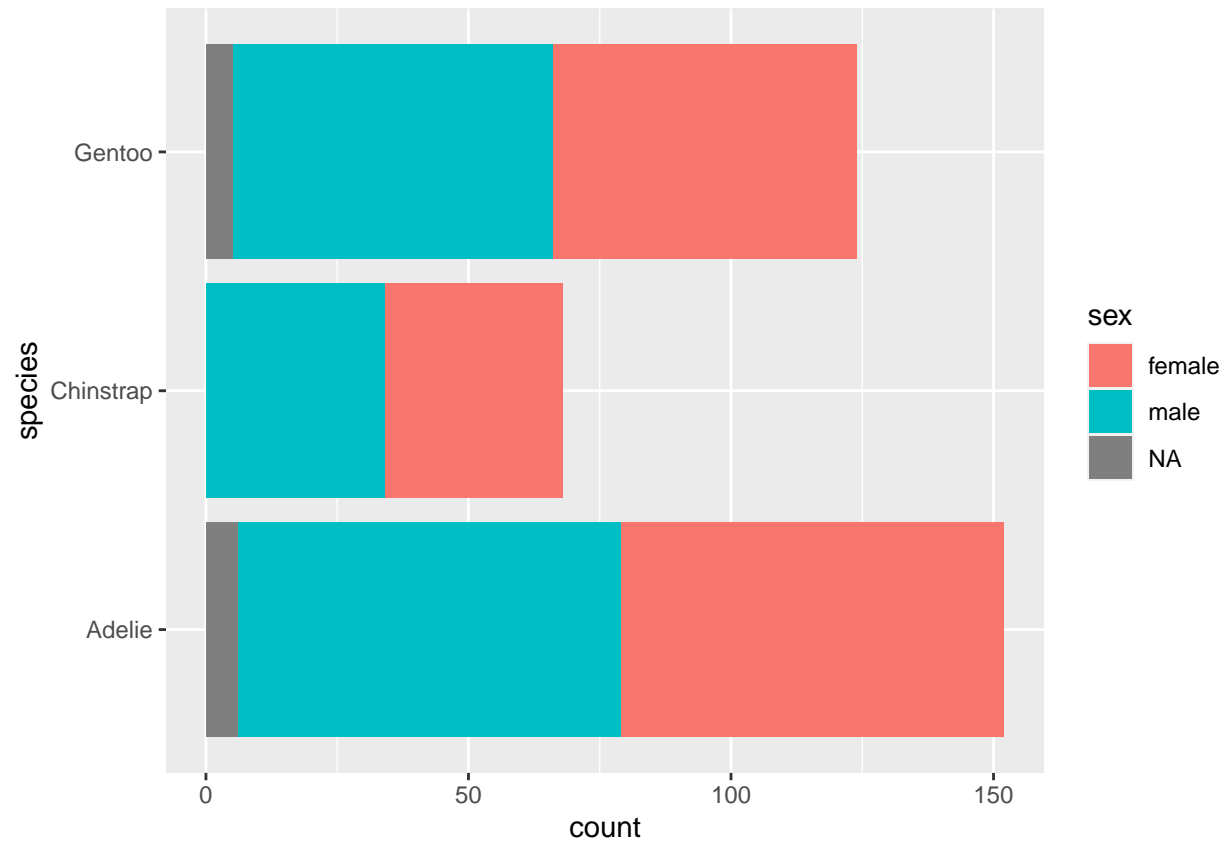
```
## Warning: Removed 2 rows containing missing values (geom_point).
```



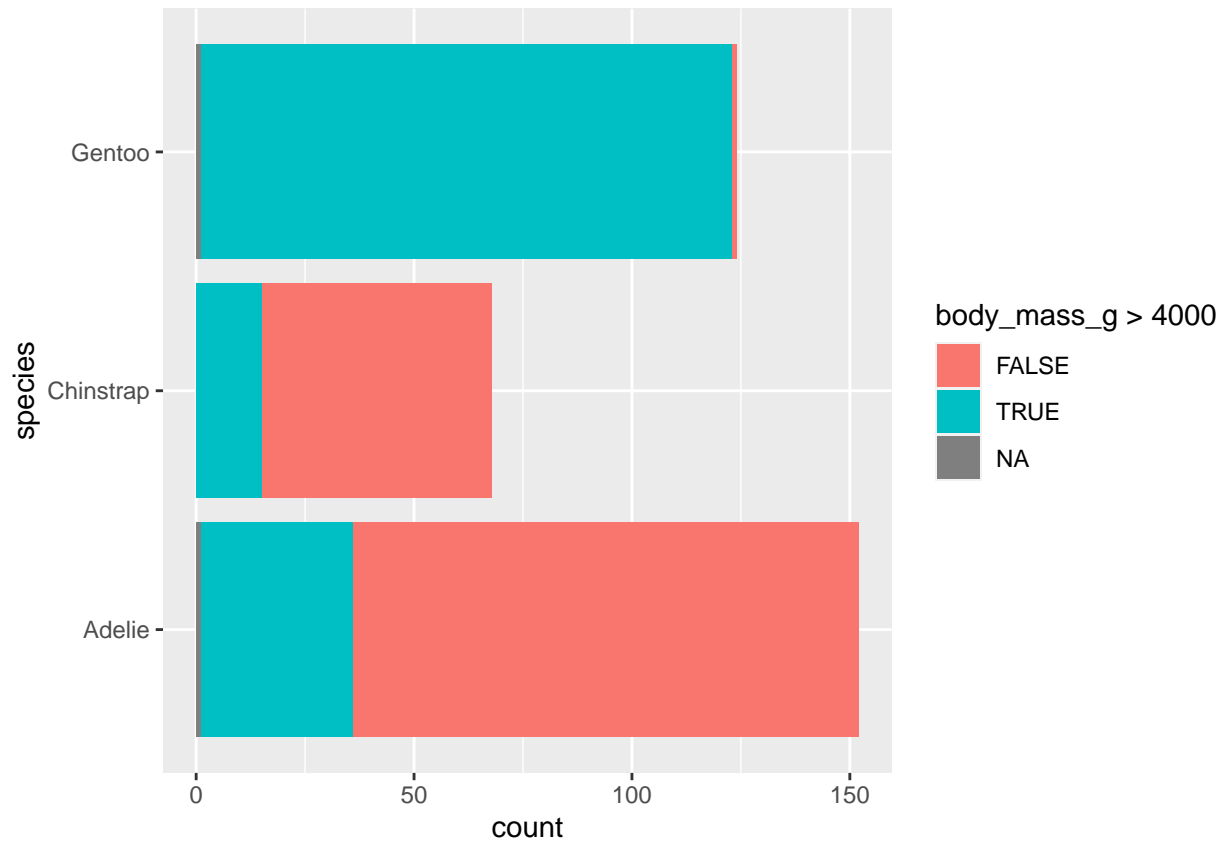
```
penguins %>%
  ggplot(aes(y = species)) +
    geom_bar(fill = 'grey', colour = 'red')
```



```
# See the count of species  
penguins %>%  
  ggplot() +  
  geom_bar(aes(y = species, fill = sex))
```



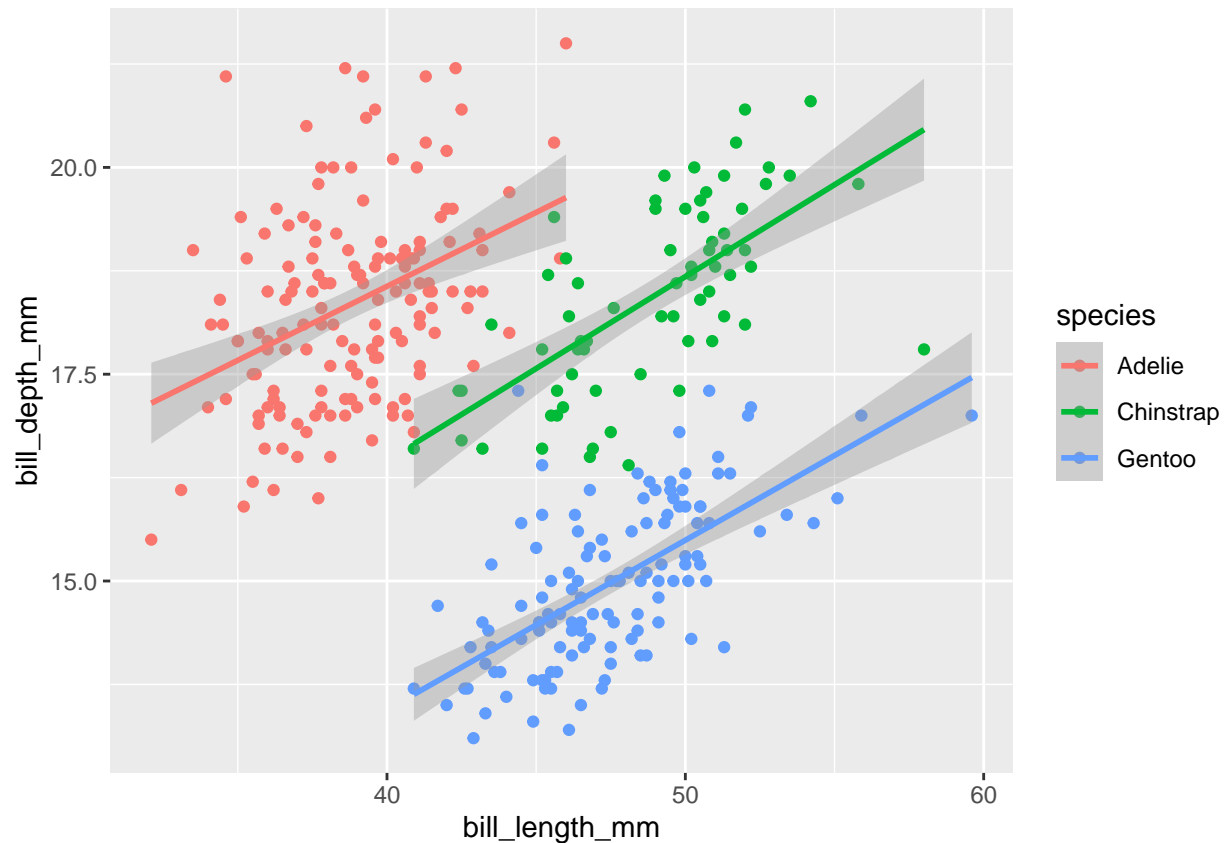
```
# Let's look at body mass of penguin species
penguins %>%
  ggplot() +
  geom_bar(aes(y = species,
               fill = body_mass_g > 4000))
```



```
# let's look at the relation between bill length and bill depth for different
# penguin species
penguins %>%
  ggplot(aes(x = bill_length_mm,
             y = bill_depth_mm,
             colour = species)) +
  geom_point() +
  geom_smooth(method = 'lm', formula = 'y ~ x')
```

```
## Warning: Removed 2 rows containing non-finite values (stat_smooth).
```

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

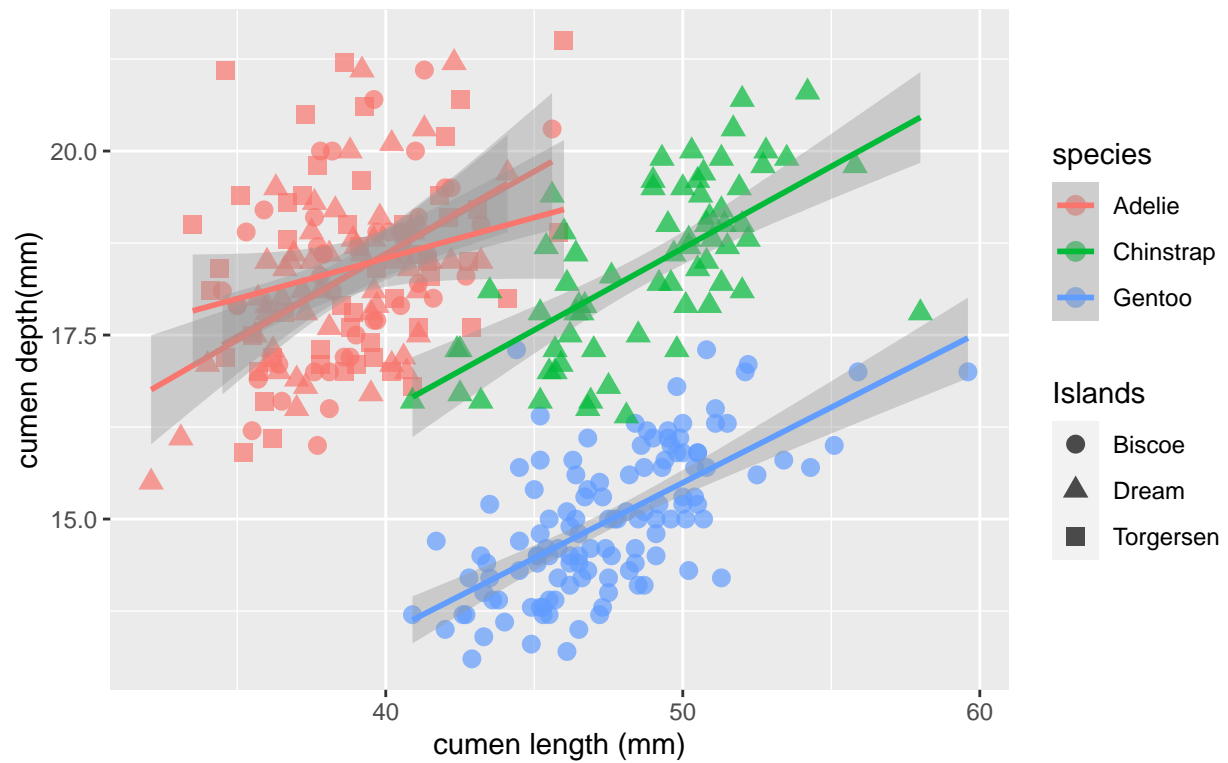


```
# let's look at the relation between bill length and bill depth for different
# penguin species and islands
penguins %>%
  ggplot(aes(x = bill_length_mm,
             y = bill_depth_mm,
             colour = species,
             shape = island)) +
  geom_point(size = 3, alpha = 0.7) +
  geom_smooth(method = 'lm', formula = 'y ~ x') +
  labs(title = 'Bill ration of palmer penguins',
       caption = 'Horst AM, Hill AP, Gorman KB (2020)',
       shape = 'Islands',
       x = 'cumen length (mm)',
       y = 'cumen depth(mm)')
```

```
## Warning: Removed 2 rows containing non-finite values (stat_smooth).
```

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

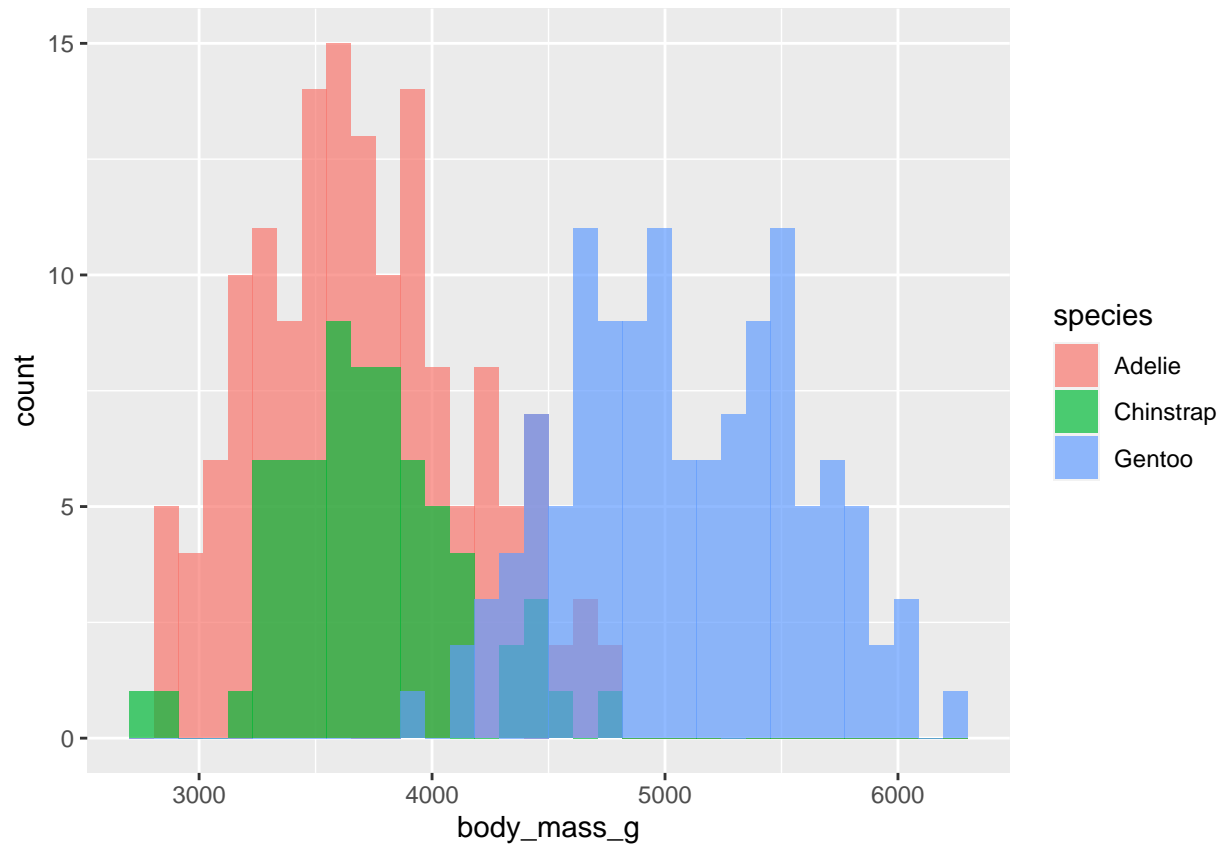
Bill ration of palmer penguins



Horst AM, Hill AP, Gorman KB (2020)

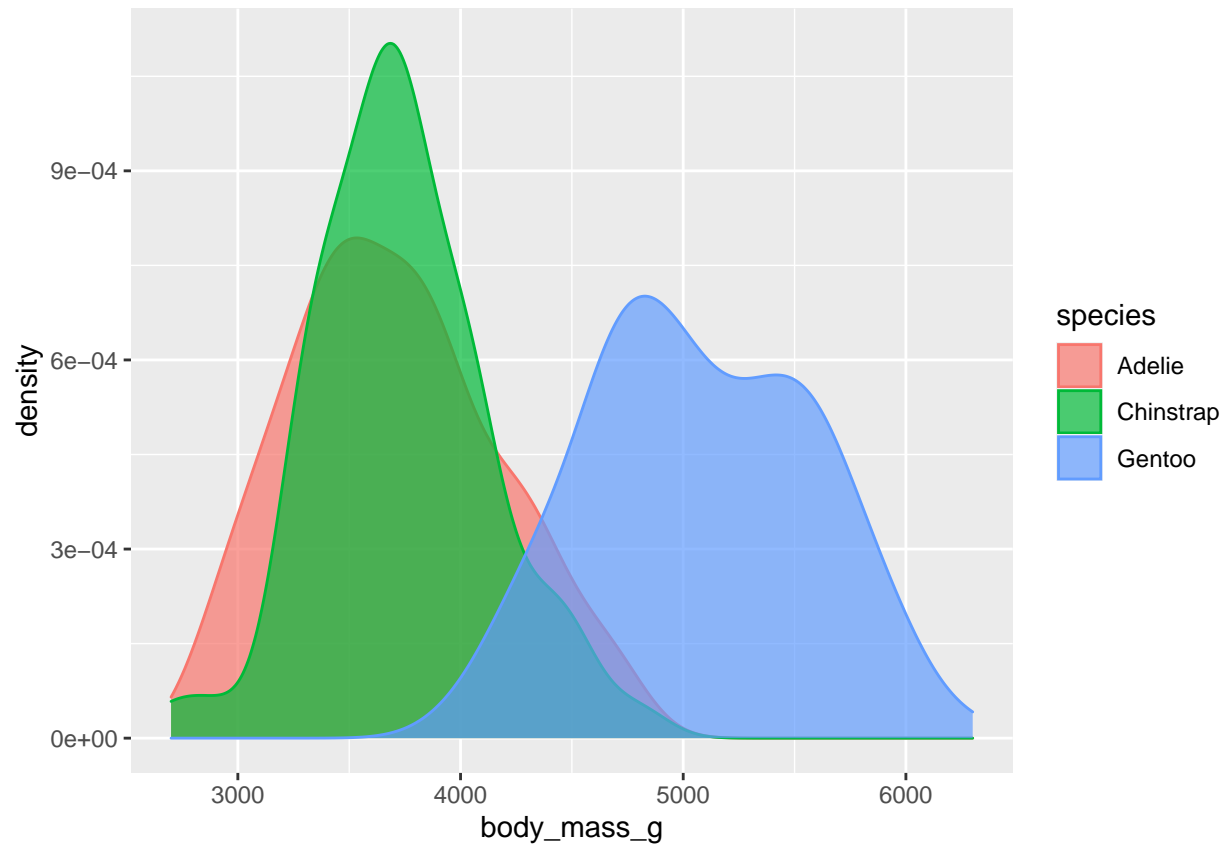
```
# Let's look at histogram of body mass
penguins %>%
  ggplot(aes(x = body_mass_g,
             fill = species)) +
  geom_histogram(bins = 35,
                alpha = 0.7,
                position = 'identity')
```

```
## Warning: Removed 2 rows containing non-finite values (stat_bin).
```

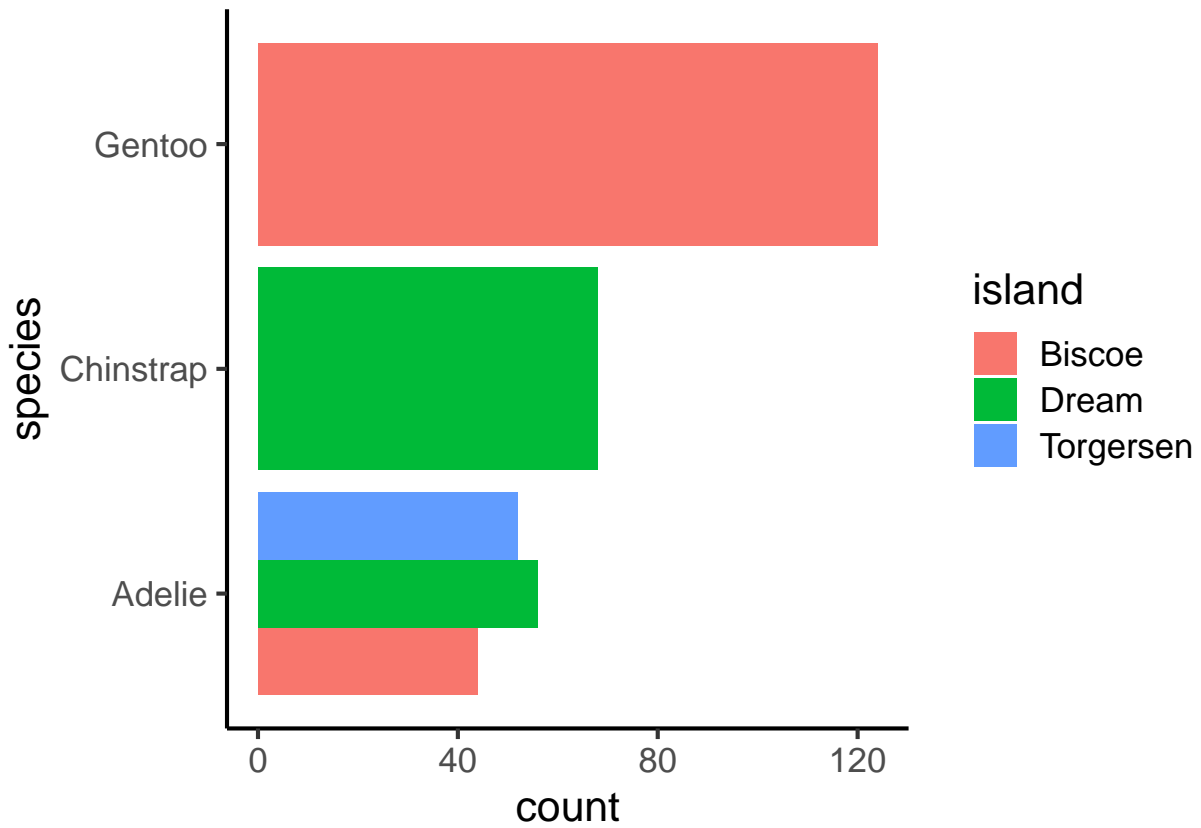


```
penguins %>%  
  ggplot(aes(x = body_mass_g,  
             fill = species,  
             colour = species)) +  
  geom_density(alpha=0.7)
```

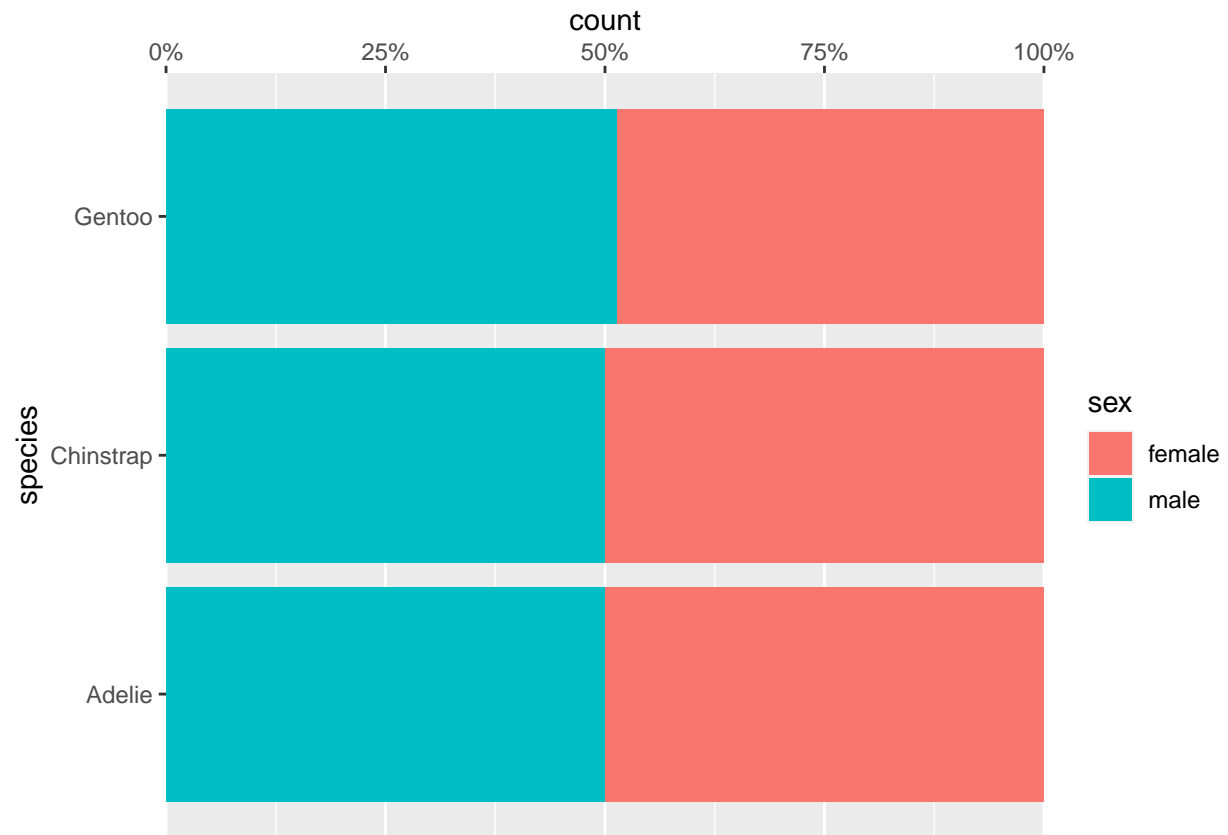
```
## Warning: Removed 2 rows containing non-finite values (stat_density).
```

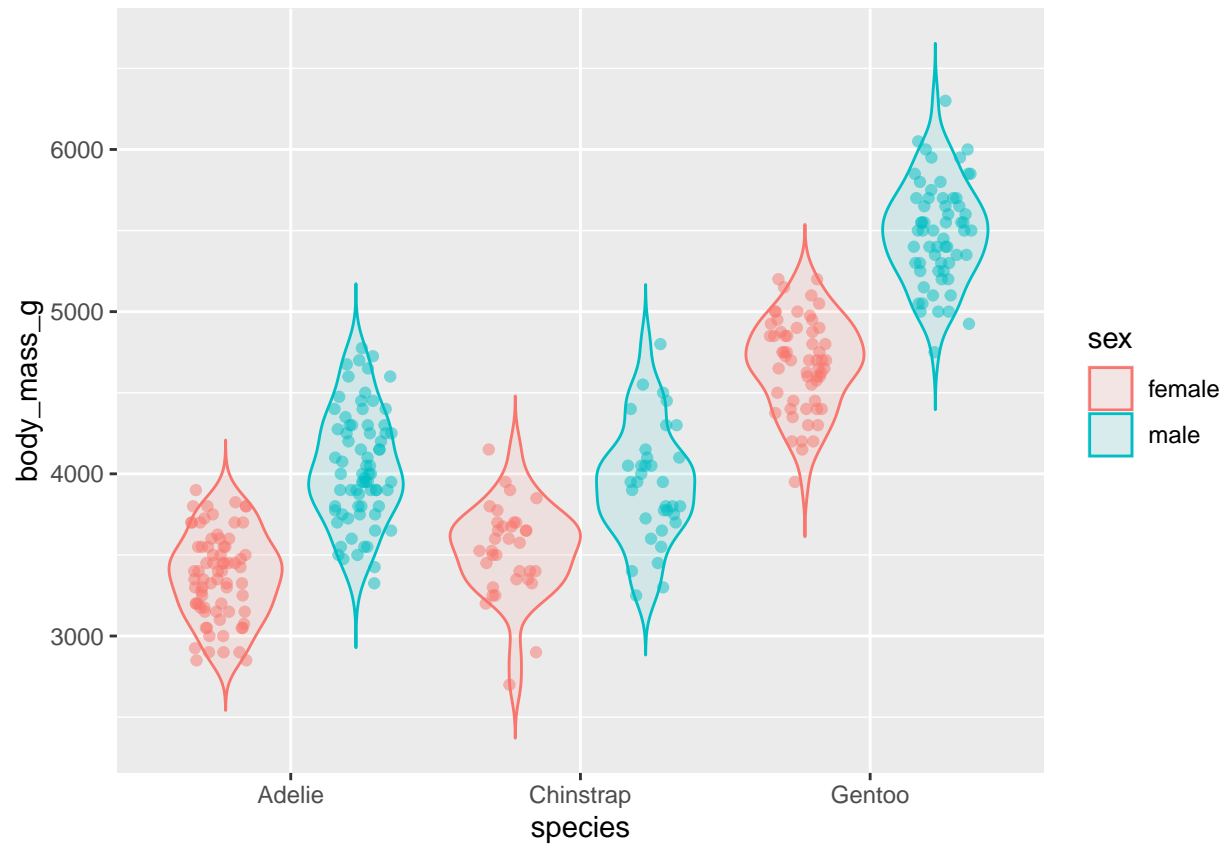
```
# Penguin species by island
penguins %>%
  ggplot() +
    geom_bar(aes(y = species,
                 fill = island),
             position = 'dodge') +
    theme_classic(16)
```



```
penguins %>%
  drop_na(sex) %>%
  ggplot() +
  geom_bar(aes(y = species,
               fill = sex),
           position = 'fill') +
  scale_x_continuous(labels = scales::percent,
                    position = 'top',
                    expand = c(0,0))
```

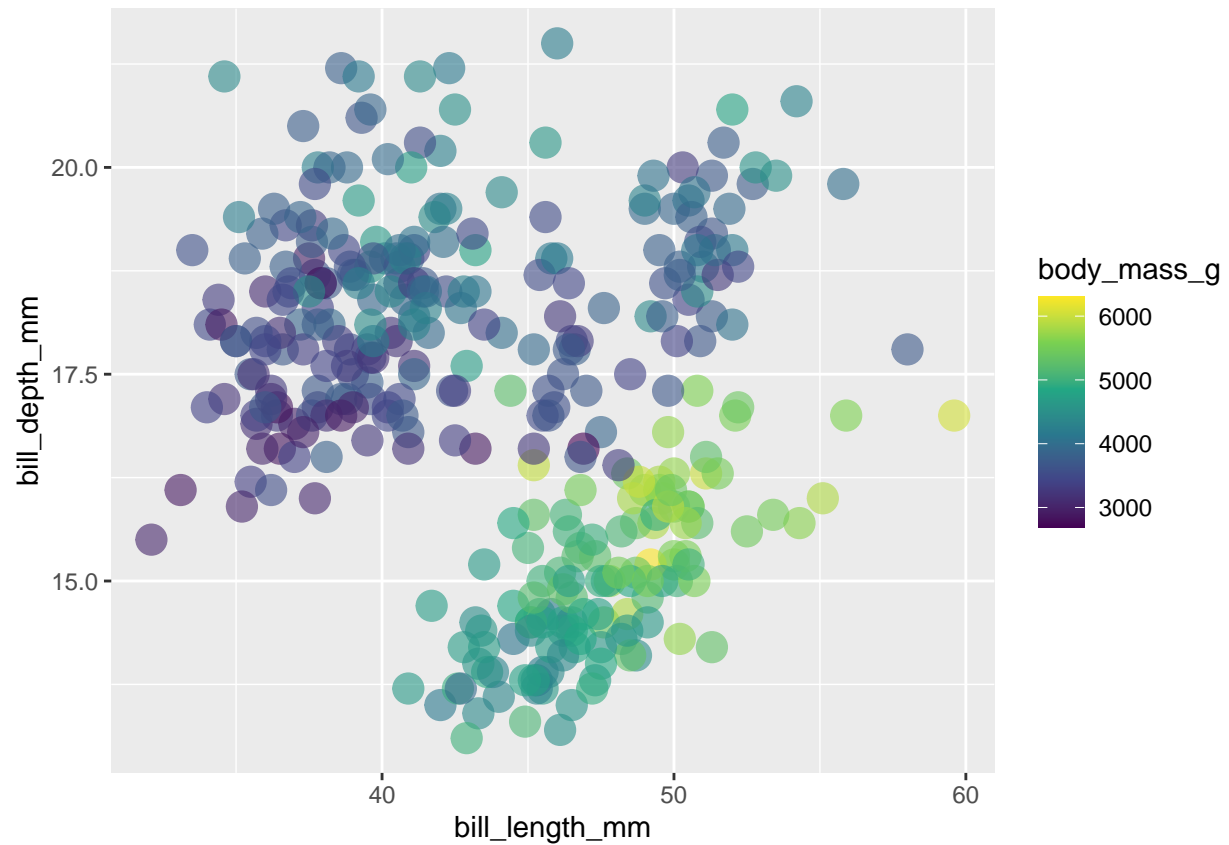


```
penguins %>%
  drop_na(sex) %>%
  ggplot(aes(x = species,
             y = body_mass_g,
             fill = sex,
             colour = sex)) +
  geom_violin(alpha = 0.1, trim = FALSE) +
  geom_point(position = position_jitterdodge(dodge.width = 1),
            alpha = 0.5,
            show.legend = F)
```



```
penguins %>%
  ggplot(aes(x = bill_length_mm,
             y = bill_depth_mm,
             colour = body_mass_g)) +
  geom_point(alpha = 0.6, size = 5) +
  scale_color_viridis_c()
```

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

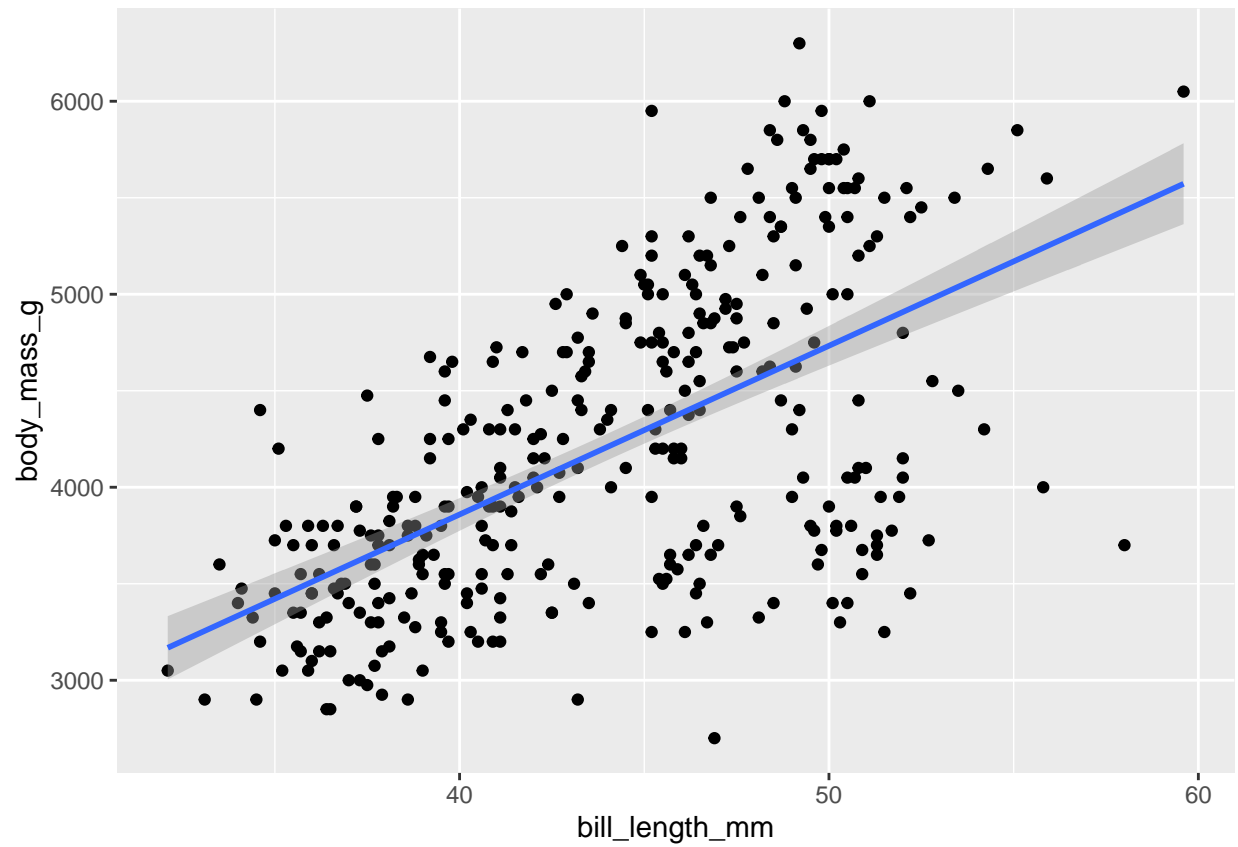


```
penguins %>%  
  ggplot(aes(x = bill_length_mm,  
             y = body_mass_g)) +  
  geom_point() +  
  geom_smooth(method = "lm")
```

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

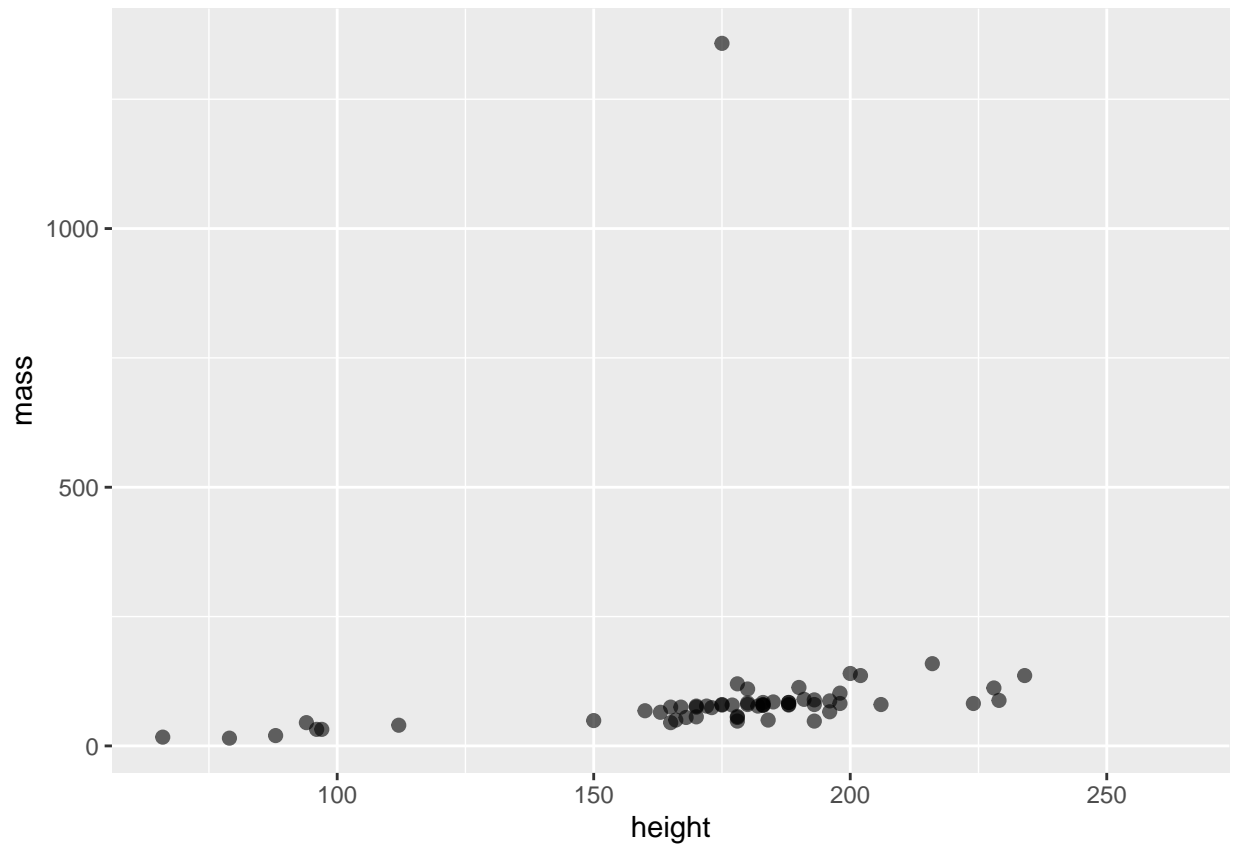
```
## Warning: Removed 2 rows containing non-finite values (stat_smooth).
```

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

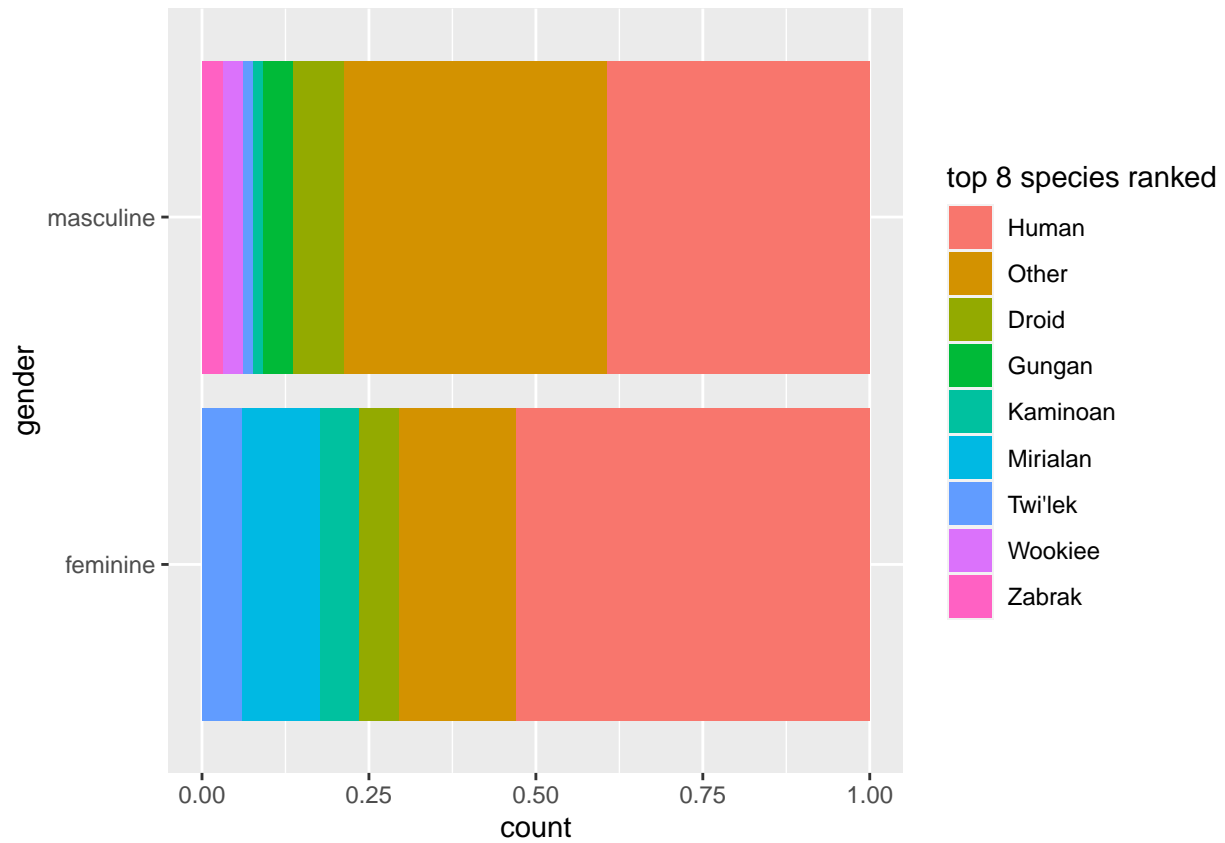


```
starwars %>%  
  ggplot(aes(x = height,  
             y = mass)) +  
  geom_point(size=2, alpha=0.6)
```

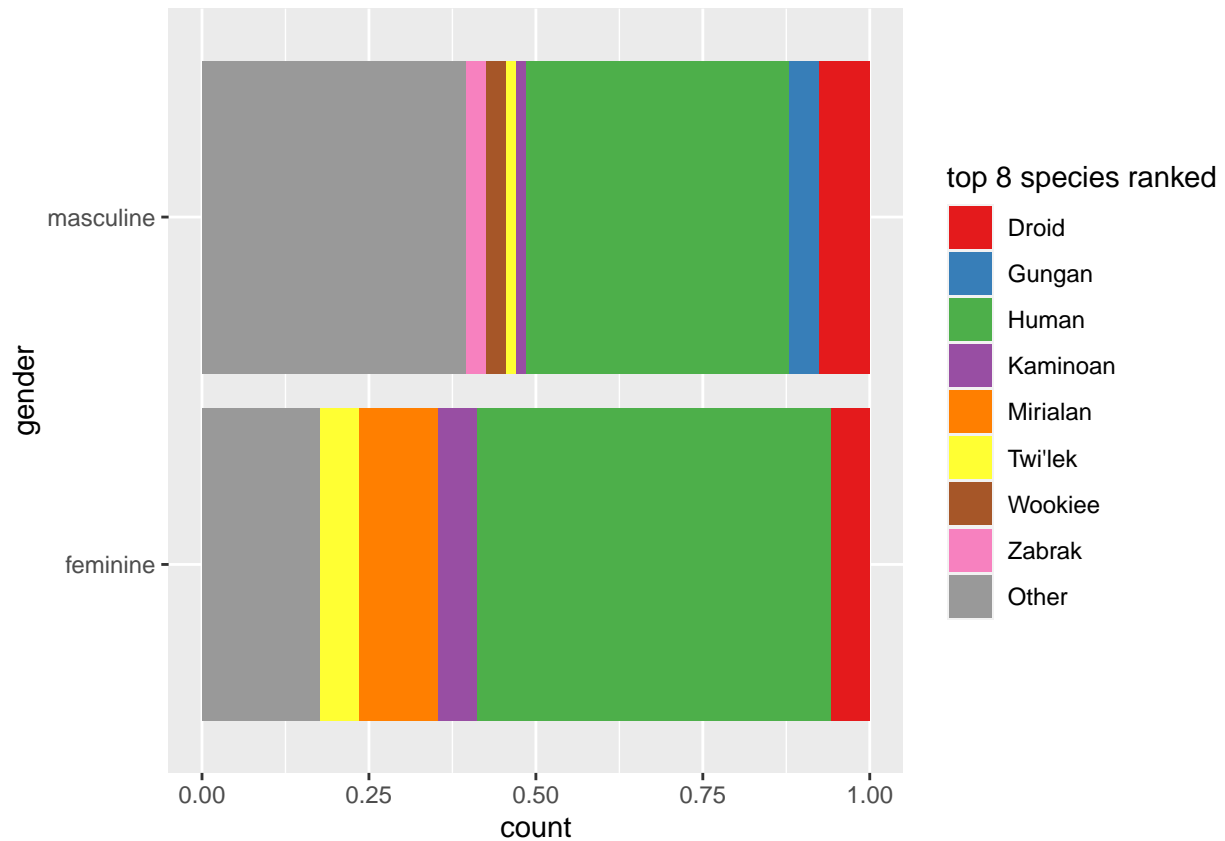
```
## Warning: Removed 28 rows containing missing values (geom_point).
```



```
# Top 8 species according to gender in star wars
filter(starwars, !is.na(gender)) %>%
  ggplot(aes(y = gender,
             fill = fct_lump_n(species, 5) %>%
               fct_infreq())) +
  geom_bar(position = 'fill') +
  labs(fill = 'top 8 species ranked') +
  scale_fill_hue()
```

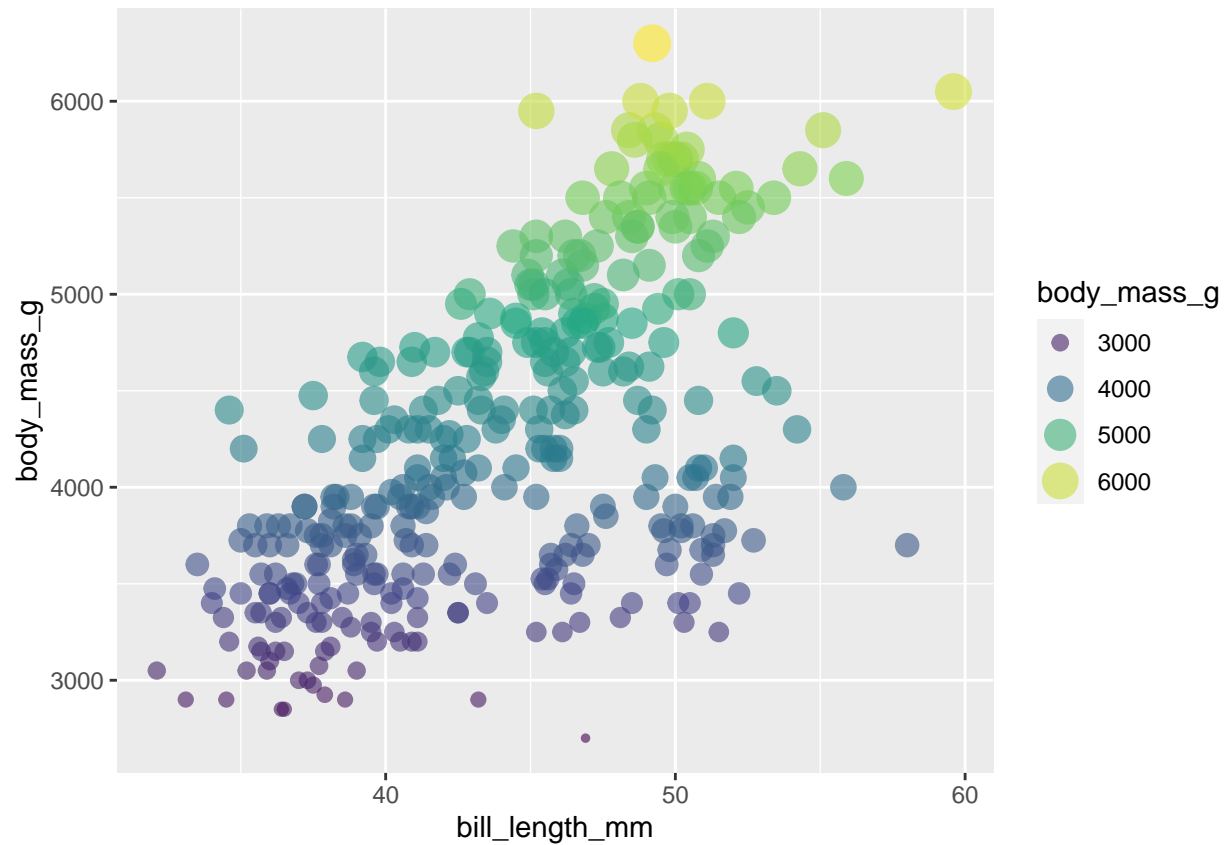


```
# Top 8 species according to gender in star wars - out of order
filter(starwars, !is.na(gender)) %>%
  ggplot(aes(y = gender,
             fill = fct_lump_n(species, 5))) +
  geom_bar(position = 'fill') +
  labs(fill = 'top 8 species ranked') +
  scale_fill_brewer(palette = "Set1")
```

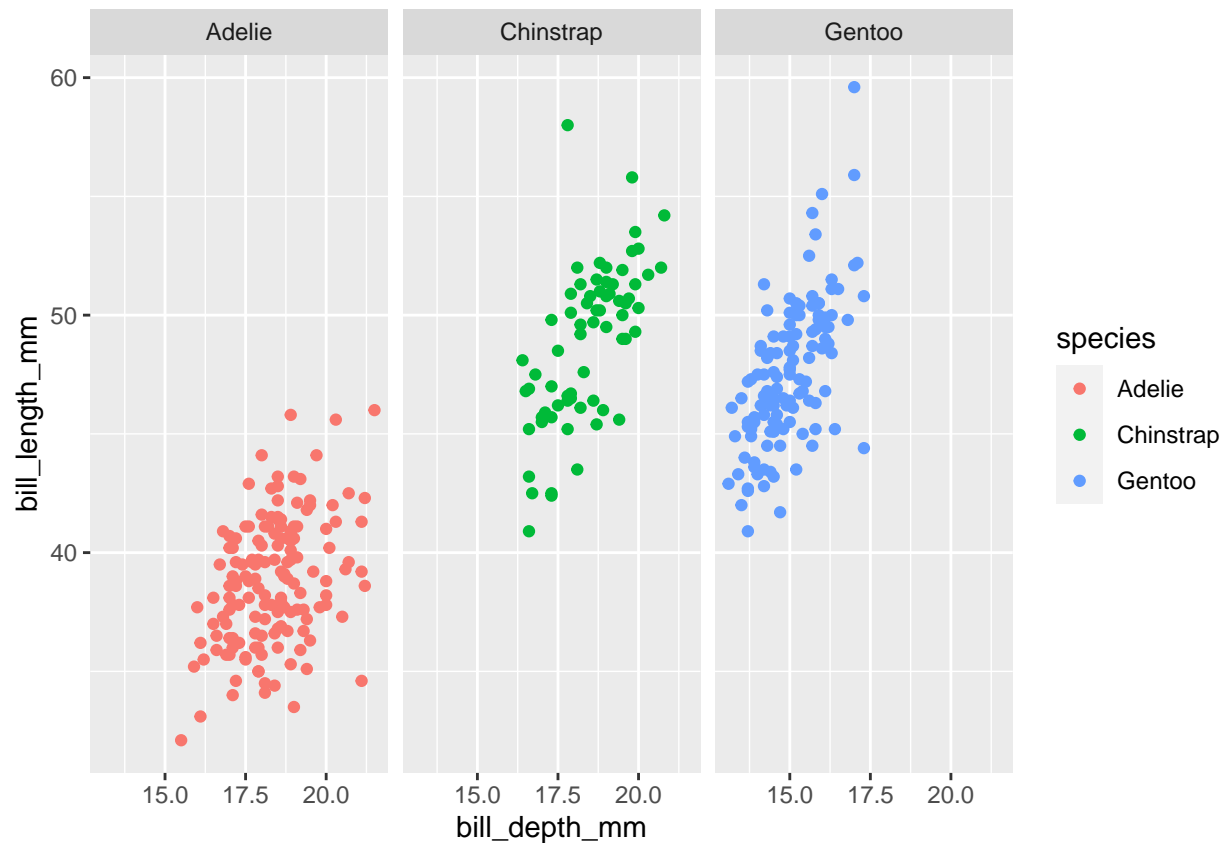
```
# Mapping size and color
penguins %>%
  ggplot(aes(x=bill_length_mm,
             y=body_mass_g,
             color=body_mass_g,
             size=body_mass_g)) +
  geom_point(alpha=0.6) +
  scale_color_viridis_c() +
  guides(colour = 'legend')
```

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



```
# Visualizing with multiple grids
penguins %>%
  ggplot(aes(x=bill_depth_mm,
             y=bill_length_mm,
             color = species)) +
  geom_point()+
  facet_wrap(~species, ncol = 3)
```

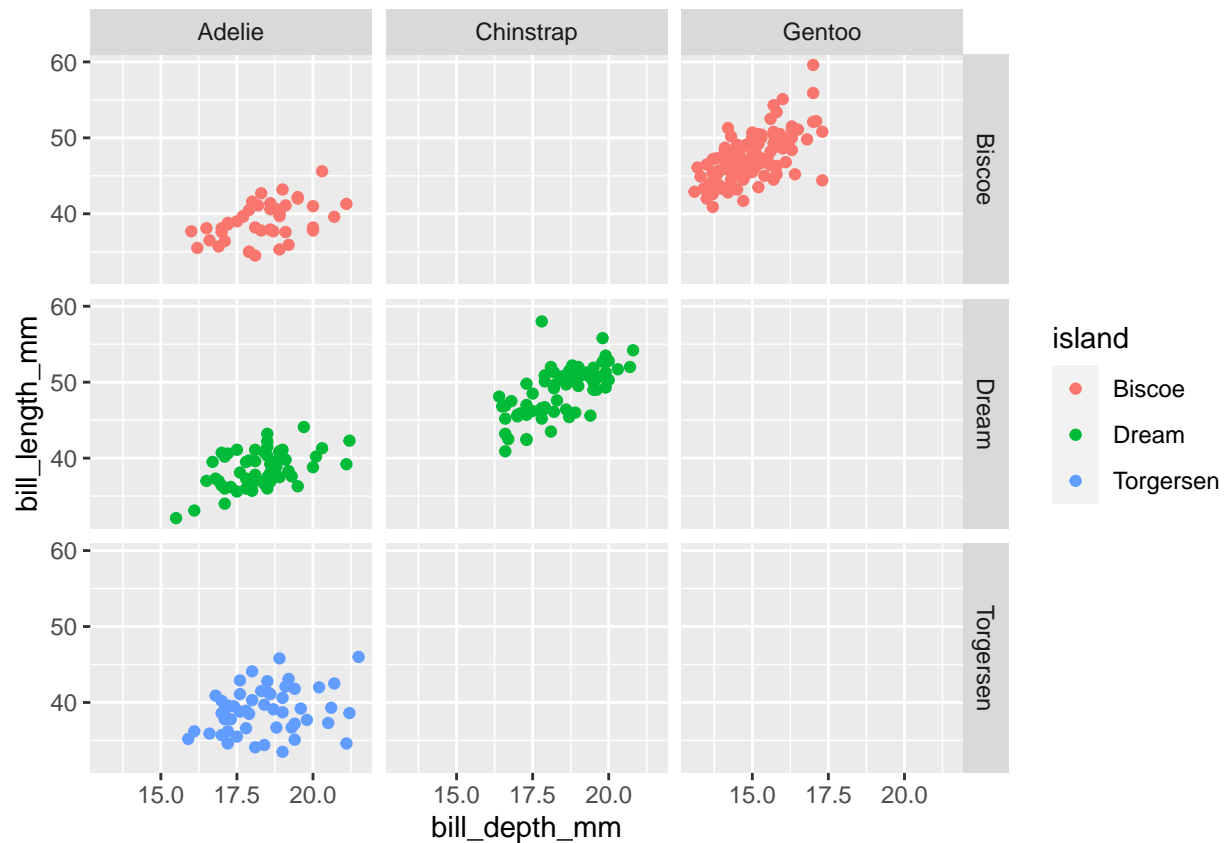
```
## Warning: Removed 2 rows containing missing values (geom_point).
```



```
penguins %>%
  ggplot(aes(x=bill_depth_mm,
             y=bill_length_mm,
             color = island,
             color = species)) +
  geom_point()+
  facet_grid(island~species)
```

```
## Warning: Duplicated aesthetics after name standardisation: colour
```

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



```
penguins %>%
  ggplot(aes(x=bill_depth_mm,
             y=bill_length_mm,
             colour = species)) +
  geom_point() +
  facet_grid(island~.) +
  labs(x='length (mm)', y = 'depth (mm)',
       title = 'Palmer Penguins',
       subtitle = 'Bill dimensions over location and species') +
  scale_fill_ipsum()
```

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

Palmer Penguins

Bill dimensions over location and species

