Penguin 1.1

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.0      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

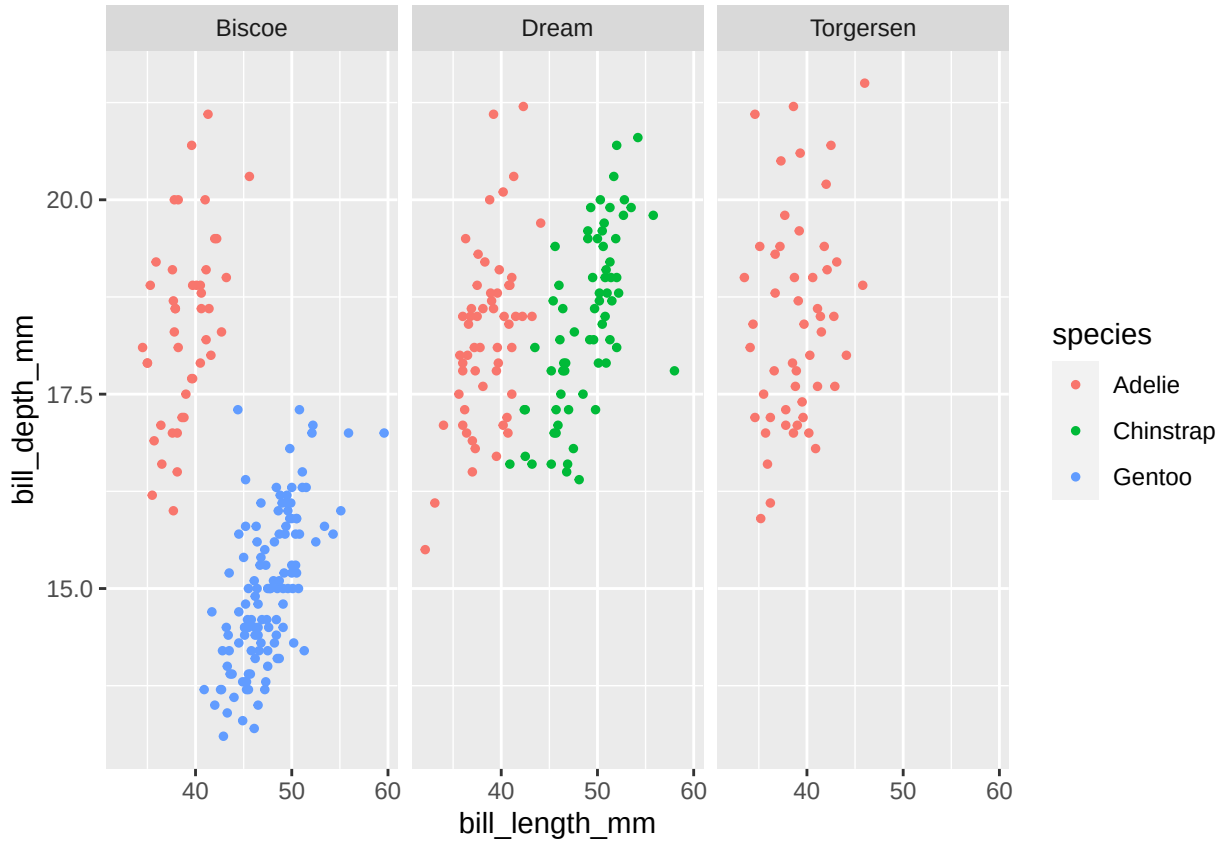
```
penguins <- read_csv("https://uwmadison.box.com/shared/static/ijh7iipc9ect1jf0z8qa2n3j7dgem1gh.csv")
```

```
## Rows: 344 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): species, island, sex
## dbl (5): bill_length_mm, bill_depth_mm, flipper_length_mm, body_mass_g, year
##
## 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,2)
```

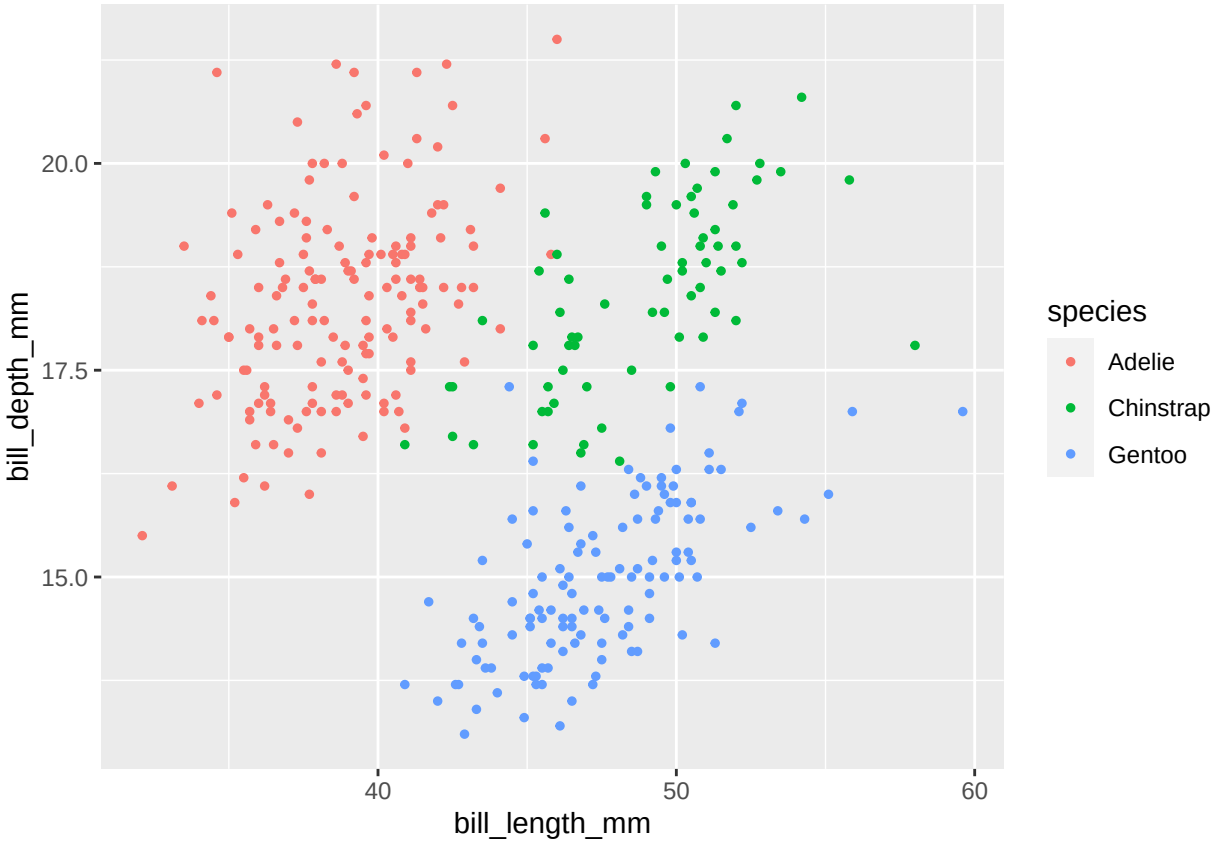
```
## # A tibble: 2 x 8
##   species island    bill_length_mm bill_depth_mm flipper_l~1 body_~2 sex    year
##   <chr>   <chr>              <dbl>         <dbl>       <dbl>   <dbl> <chr> <dbl>
## 1 Adelie  Torgersen           39.1          18.7         181    3750 male   2007
## 2 Adelie  Torgersen           39.5          17.4         186    3800 fema~  2007
## # ... with abbreviated variable names 1: flipper_length_mm, 2: body_mass_g
```

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




```
ggplot(penguins)+
  geom_point(aes(bill_length_mm, bill_depth_mm, colour = species),size = 1)
```

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



Simpson's paradox: It is a phenomenon in probability and statistics in which a trend appears in several groups of data but disappears or reverses when the groups are combined In the first plot above, every island has species Adelie, but when combining the plots together, the trend in each island disappeared. This phemomeonon satisfied the content of Simpson's paradox. The Adelie specie is around 50%-100% in each plot in the first graph; however, when comes to the second plot, the Adelie specie's porprotion is no longer more than 60%(visually)