LCBC

LIFESPAN CHANGES
in brain and cognition

Tidy data wrangling

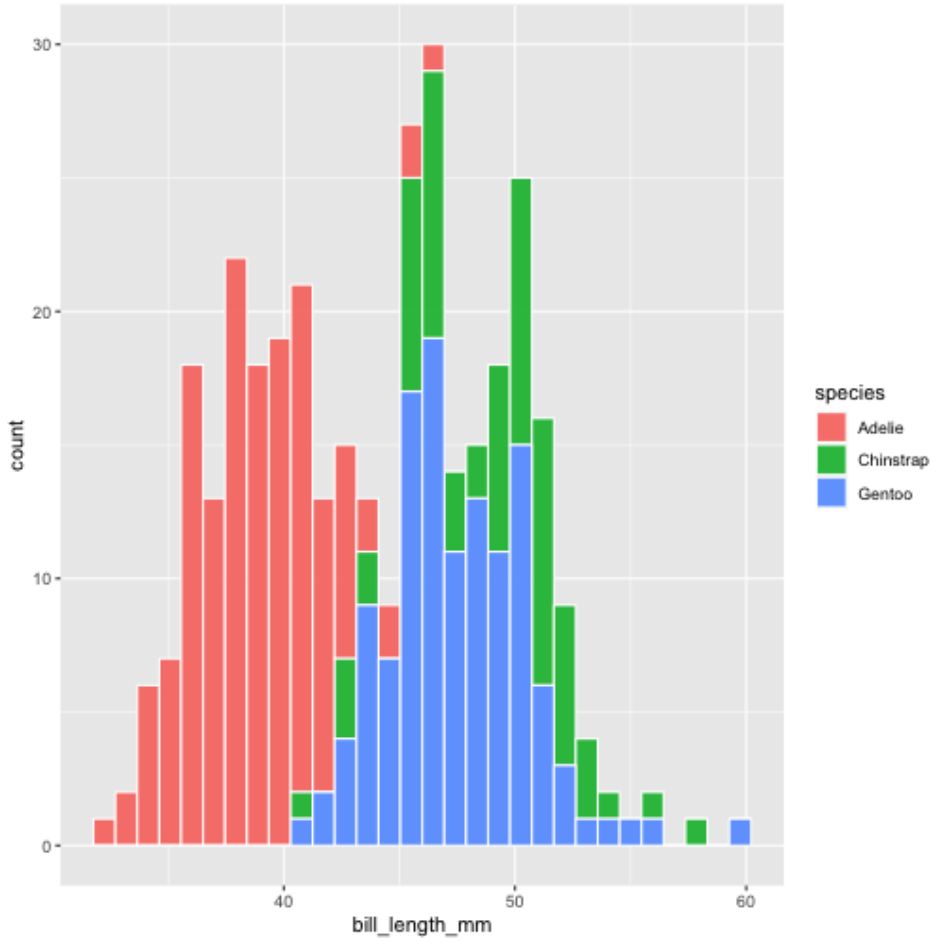
demonstration

Athanasia Monika Mowinckel

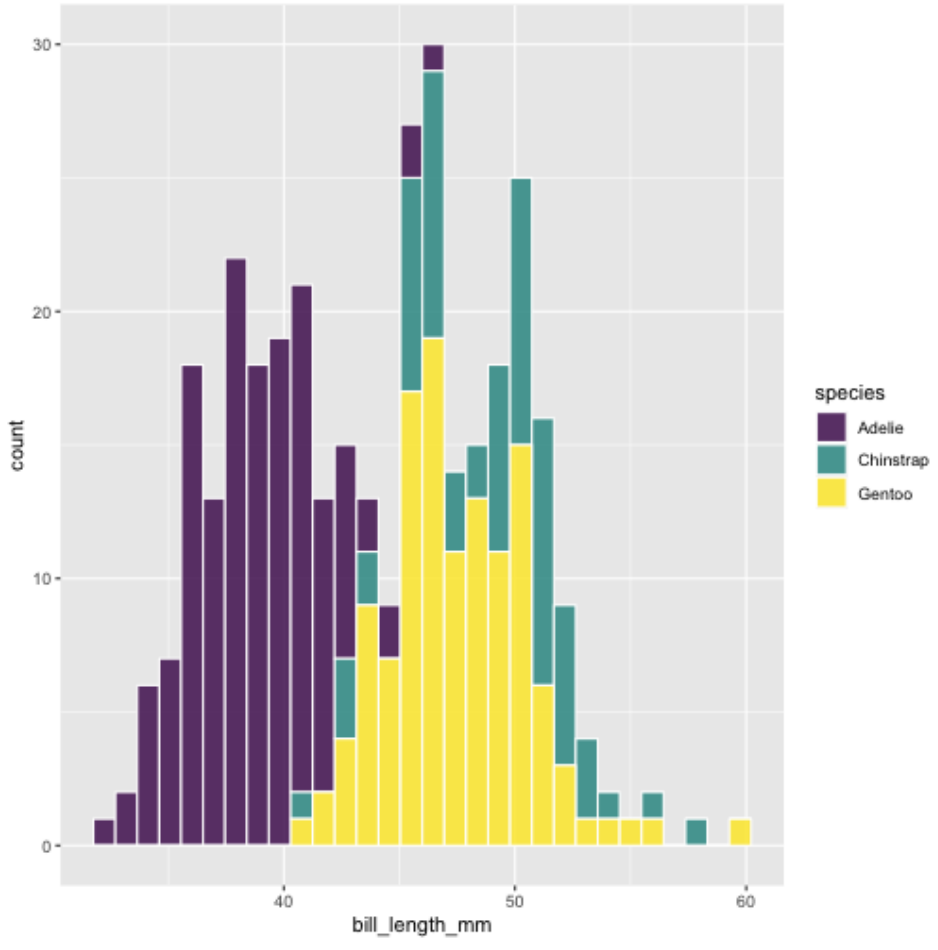
Sept. 15th 2020

ggplot2

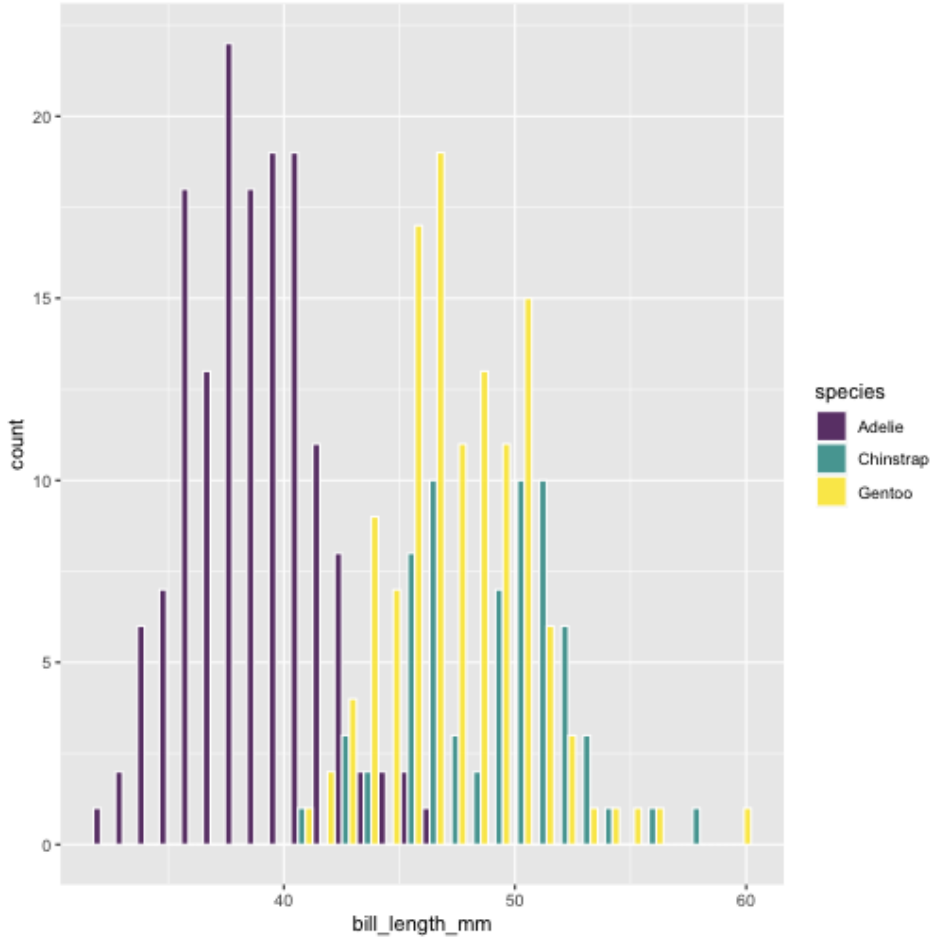
```r
ggplot(penguins,
       aes(x = bill_length_mm,
           fill = species)) +
  geom_histogram(alpha = .9,
                 colour = "white")
```



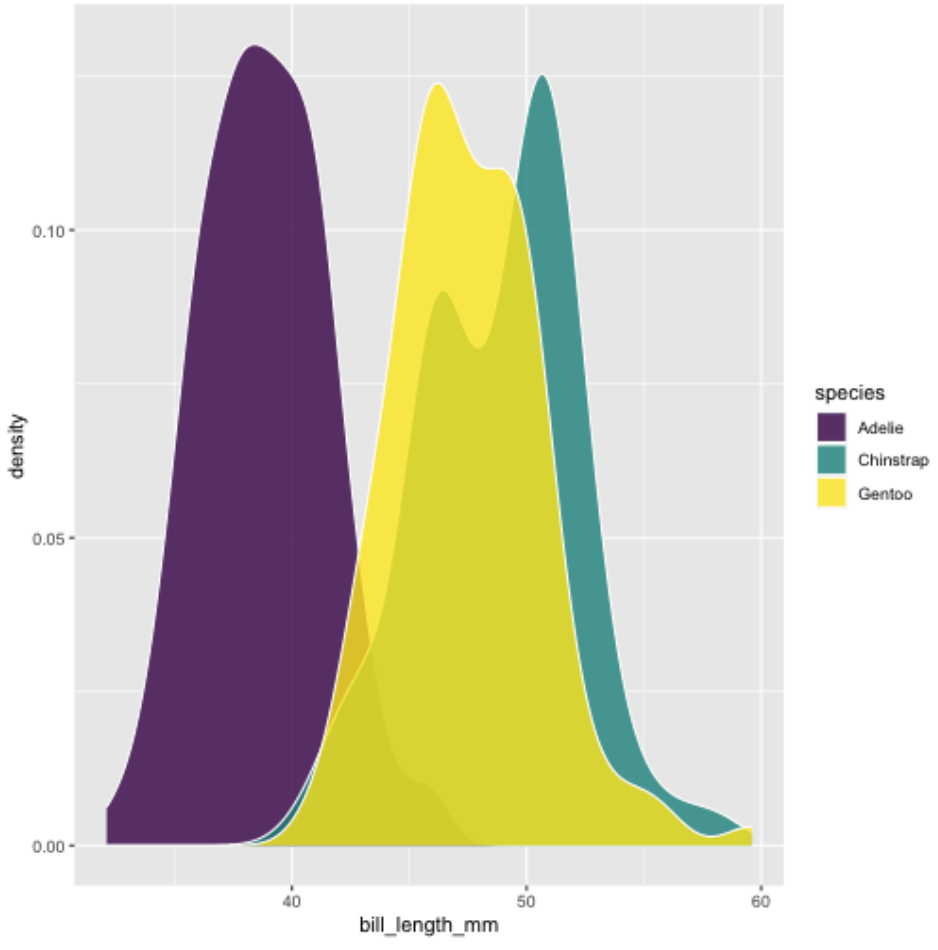
```
ggplot(penguins,
       aes(x = bill_length_mm,
           fill = species)) +
  geom_histogram(alpha = .8,
                 colour = "white") +
  scale_fill_viridis_d()
```



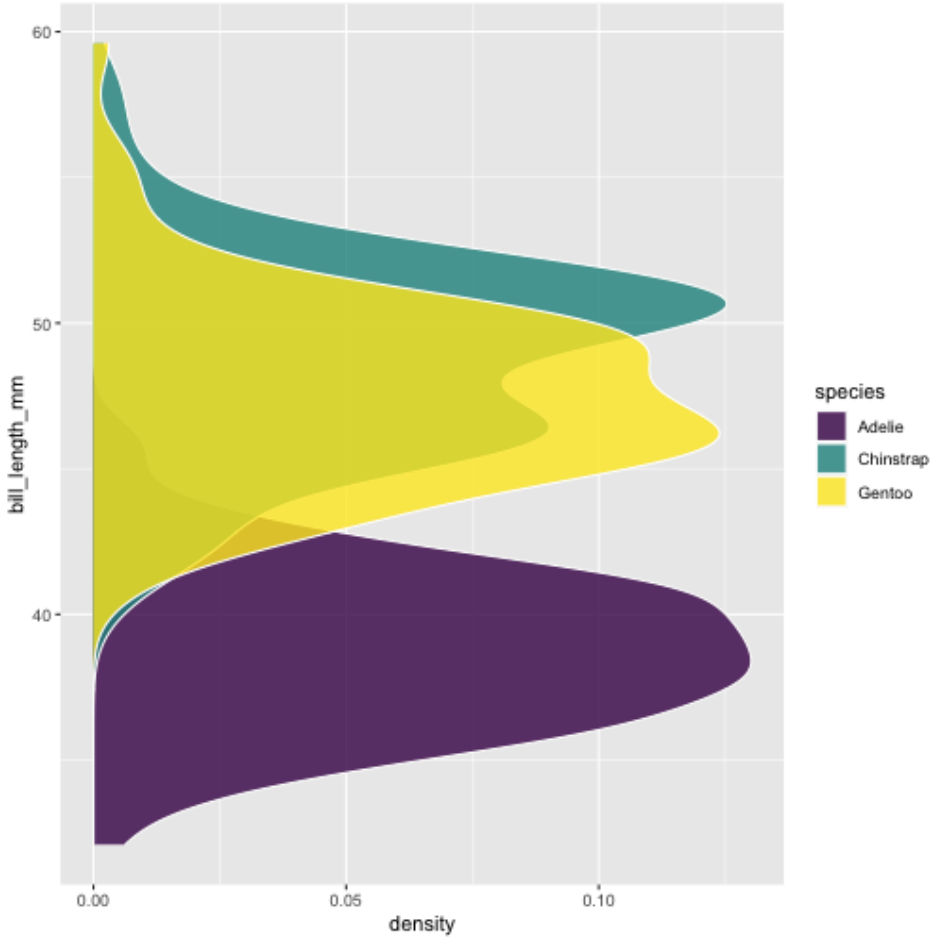
```
ggplot(penguins,
       aes(x = bill_length_mm,
           fill = species)) +
  geom_histogram(alpha = .8,
                 colour = "white",
                 position = "dodge") +
  scale_fill_viridis_d()
```



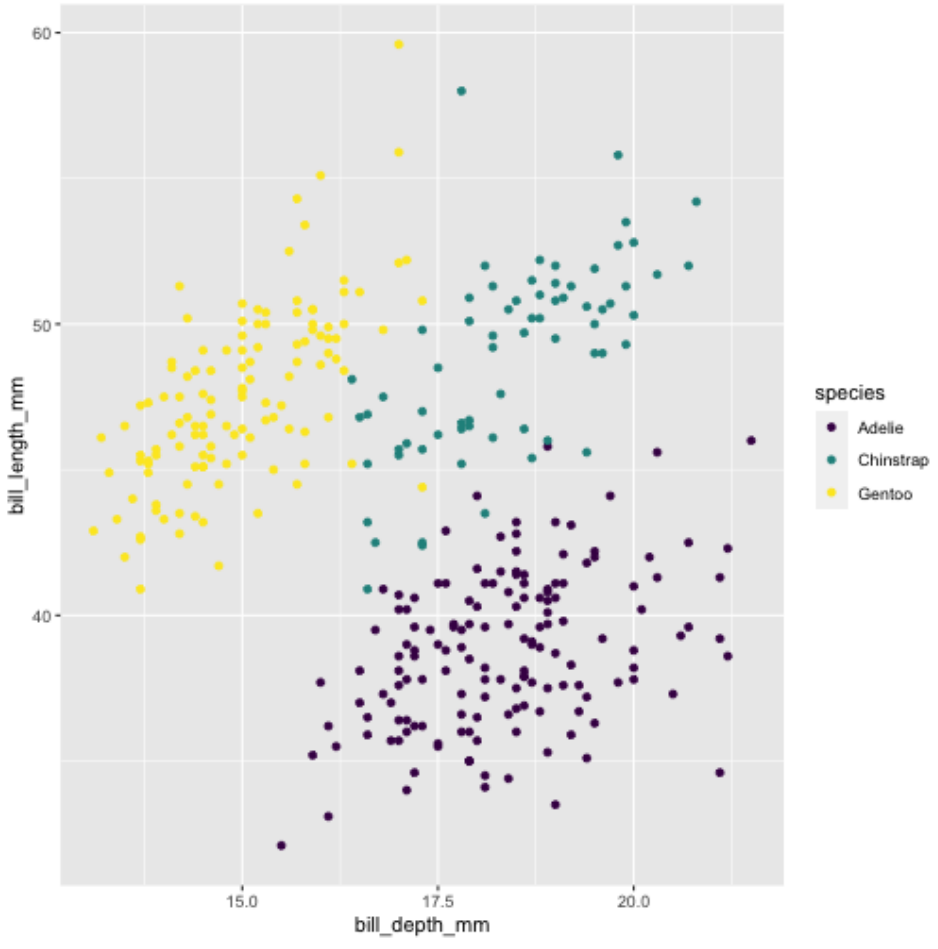
```
ggplot(penguins,
       aes(x = bill_length_mm,
           fill = species)) +
  geom_density(
    alpha = .8,
    colour = "white",
    position = "dodge") +
  scale_fill_viridis_d()
```



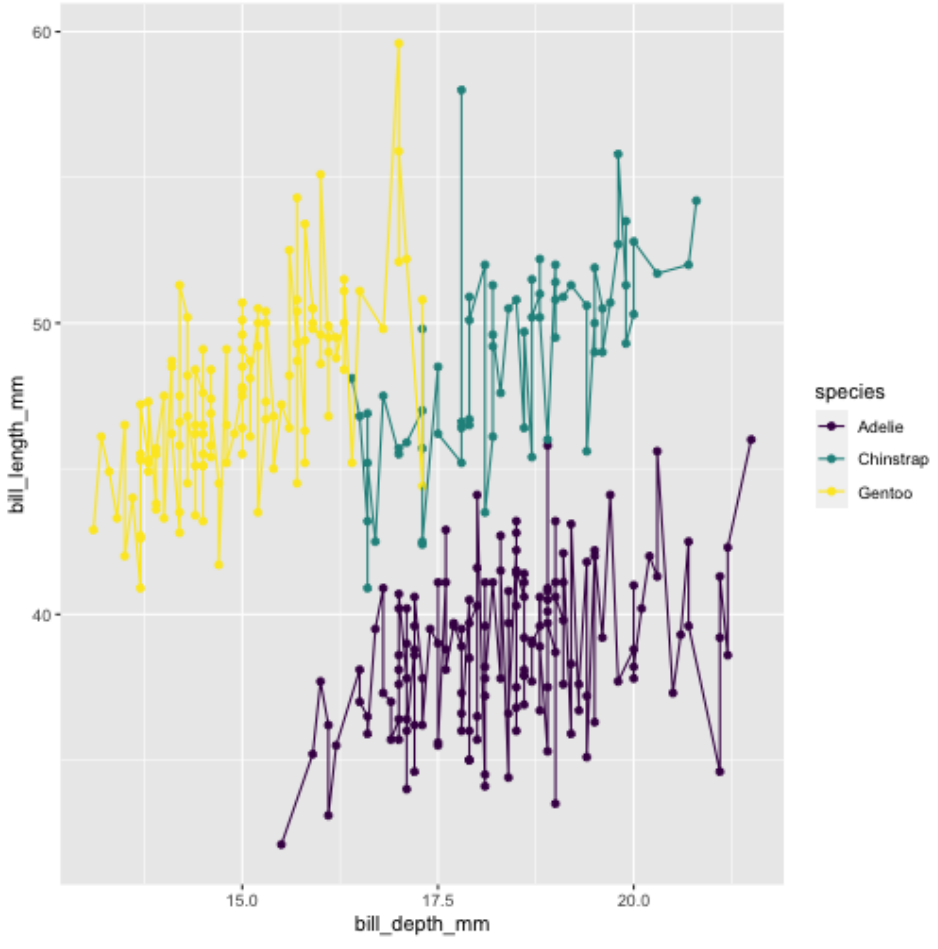
```
ggplot(penguins,
       aes(y = bill_length_mm,
           fill = species)) +
  geom_density(
    alpha = .8,
    colour = "white",
    position = "dodge") +
  scale_fill_viridis_d()
```



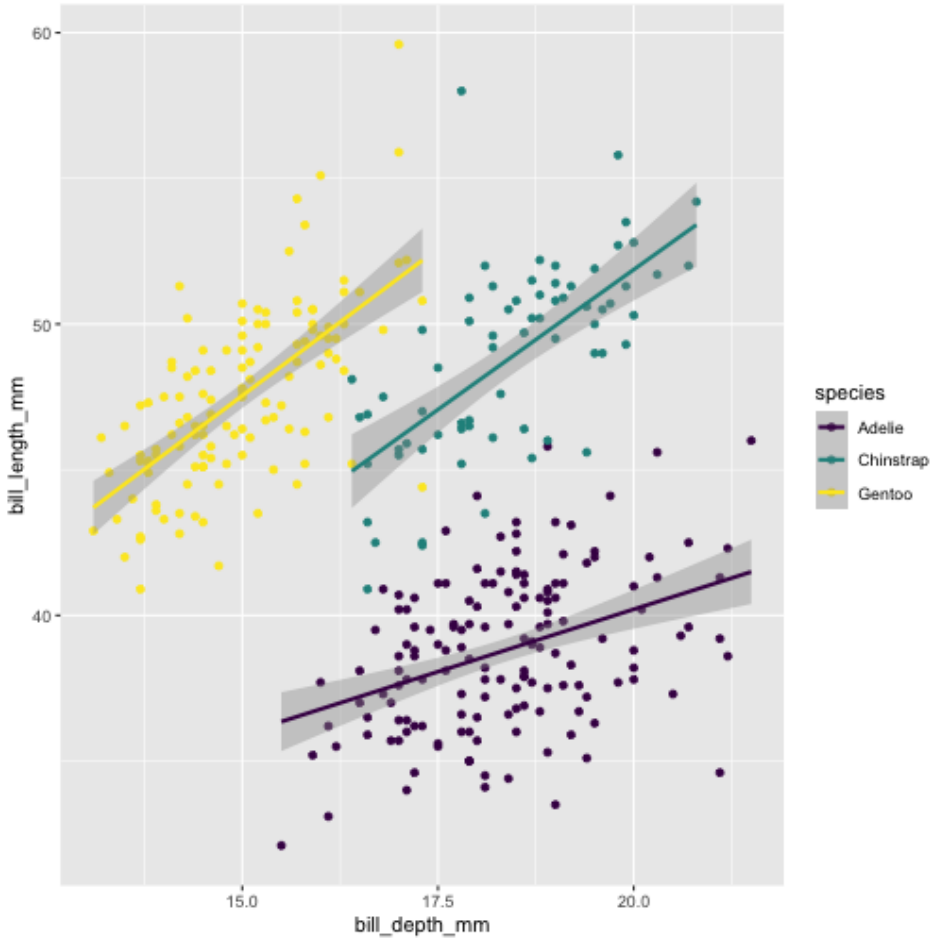
```
ggplot(penguins,
       aes(y = bill_length_mm,
           x = bill_depth_mm,
           colour = species)) +
  geom_point() +
  scale_colour_viridis_d()
```



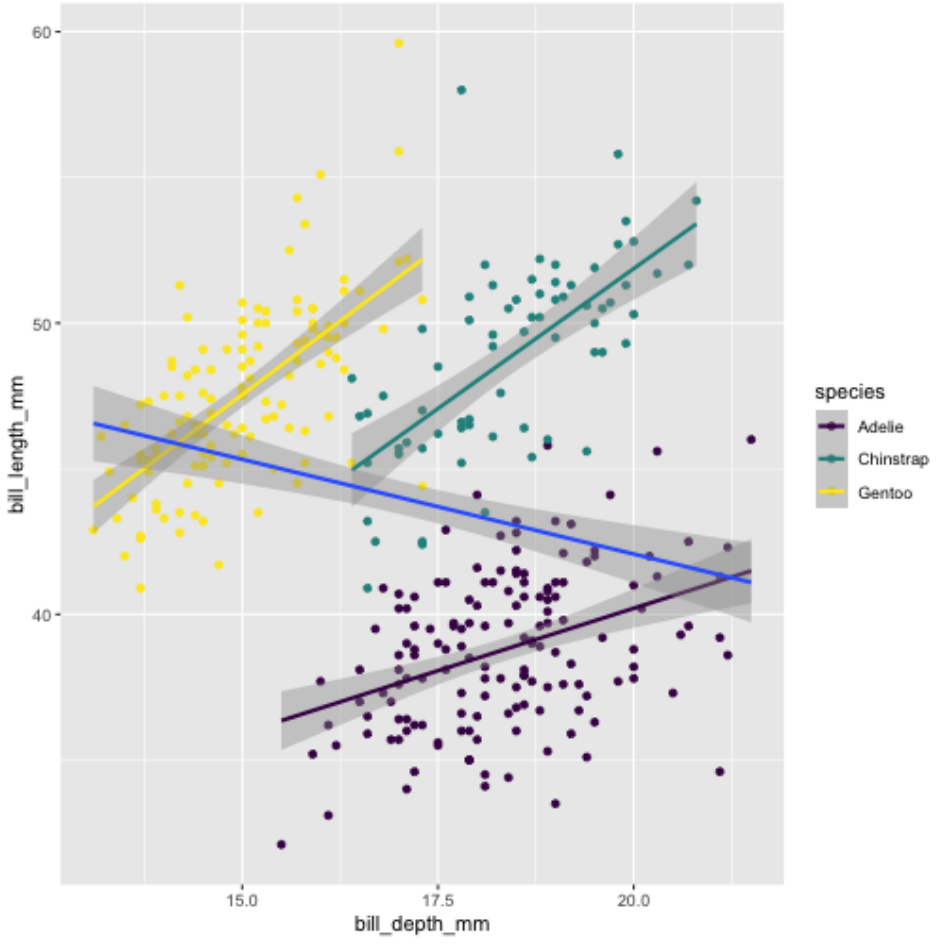
```
ggplot(penguins,
       aes(y = bill_length_mm,
           x = bill_depth_mm,
           colour = species)) +
  geom_point() +
  geom_line() +
  scale_colour_viridis_d()
```



```
ggplot(penguins,
       aes(y = bill_length_mm,
           x = bill_depth_mm,
           colour = species)) +
  geom_point() +
  geom_smooth(method = "lm") +
  scale_colour_viridis_d()
```



```
ggplot(penguins,
       aes(y = bill_length_mm,
           x = bill_depth_mm)) +
  geom_point(aes(colour = species)) +
  geom_smooth(aes(colour = species),
              method = "lm") +
  geom_smooth(method = "lm") +
  scale_colour_viridis_d()
```



filter

```
filter(penguins, species == "Chinstrap")
```

```
## # A tibble: 68 x 8
##    species island bill_length_mm bill_depth_mm flipper_length_… body_mass_g
##    <fct>   <fct>           <dbl>         <dbl>             <int>       <int>
##  1 Chinst… Dream            46.5          17.9               192        3500
##  2 Chinst… Dream            50            19.5               196        3900
##  3 Chinst… Dream            51.3          19.2               193        3650
##  4 Chinst… Dream            45.4          18.7               188        3525
##  5 Chinst… Dream            52.7          19.8               197        3725
##  6 Chinst… Dream            45.2          17.8               198        3950
##  7 Chinst… Dream            46.1          18.2               178        3250
##  8 Chinst… Dream            51.3          18.2               197        3750
##  9 Chinst… Dream            46            18.9               195        4150
## 10 Chinst… Dream            51.3          19.9               198        3700
## # … with 58 more rows, and 2 more variables: sex <fct>, year <int>
```

```
filter(penguins, island == "Dream")
```

```
## # A tibble: 124 x 8
##    species island bill_length_mm bill_depth_mm flipper_length_… body_mass_g
##    <fct>   <fct>           <dbl>         <dbl>             <int>       <int>
##  1 Adelie  Dream            39.5          16.7               178        3250
##  2 Adelie  Dream            37.2          18.1               178        3900
##  3 Adelie  Dream            39.5          17.8               188        3300
##  4 Adelie  Dream            40.9          18.9               184        3900
##  5 Adelie  Dream            36.4          17                 195        3325
##  6 Adelie  Dream            39.2          21.1               196        4150
##  7 Adelie  Dream            38.8          20                 190        3950
##  8 Adelie  Dream            42.2          18.5               180        3550
##  9 Adelie  Dream            37.6          19.3               181        3300
## 10 Adelie  Dream            39.8          19.1               184        4650
## # … with 114 more rows, and 2 more variables: sex <fct>, year <int>
```

```
filter(penguins, bill_length_mm > 40)
```

```
## # A tibble: 242 x 8
##    species island bill_length_mm bill_depth_mm flipper_length_… body_mass_g
##    <fct>   <fct>           <dbl>         <dbl>             <int>       <int>
##  1 Adelie  Torge…           40.3          18                 195        3250
##  2 Adelie  Torge…           42            20.2               190        4250
##  3 Adelie  Torge…           41.1          17.6               182        3200
##  4 Adelie  Torge…           42.5          20.7               197        4500
##  5 Adelie  Torge…           46            21.5               194        4200
##  6 Adelie  Biscoe           40.6          18.6               183        3550
##  7 Adelie  Biscoe           40.5          17.9               187        3200
##  8 Adelie  Biscoe           40.5          18.9               180        3950
##  9 Adelie  Dream            40.9          18.9               184        3900
## 10 Adelie  Dream            42.2          18.5               180        3550
## # … with 232 more rows, and 2 more variables: sex <fct>, year <int>
```

```
filter(penguins, bill_length_mm > 40, species == "Chinstrap")
```

```
## # A tibble: 68 x 8
##    species island bill_length_mm bill_depth_mm flipper_length_… body_mass_g
##    <fct>   <fct>           <dbl>         <dbl>             <int>       <int>
##  1 Chinst… Dream            46.5          17.9               192        3500
##  2 Chinst… Dream            50            19.5               196        3900
##  3 Chinst… Dream            51.3          19.2               193        3650
##  4 Chinst… Dream            45.4          18.7               188        3525
##  5 Chinst… Dream            52.7          19.8               197        3725
##  6 Chinst… Dream            45.2          17.8               198        3950
##  7 Chinst… Dream            46.1          18.2               178        3250
##  8 Chinst… Dream            51.3          18.2               197        3750
##  9 Chinst… Dream            46            18.9               195        4150
## 10 Chinst… Dream            51.3          19.9               198        3700
## # … with 58 more rows, and 2 more variables: sex <fct>, year <int>
```

select

```
select(penguins, 1, 4, 6)
```

```
## # A tibble: 344 x 3
##    species bill_depth_mm body_mass_g
##    <fct>           <dbl>       <int>
##  1 Adelie           18.7        3750
##  2 Adelie           17.4        3800
##  3 Adelie           18          3250
##  4 Adelie           NA            NA
##  5 Adelie           19.3        3450
##  6 Adelie           20.6        3650
##  7 Adelie           17.8        3625
##  8 Adelie           19.6        4675
##  9 Adelie           18.1        3475
## 10 Adelie           20.2        4250
## # … with 334 more rows
```

```
select(penguins, species, island, year, bill_length_mm)
```

```
## # A tibble: 344 x 4
##    species island     year bill_length_mm
##    <fct>   <fct>     <int>          <dbl>
##  1 Adelie  Torgersen  2007           39.1
##  2 Adelie  Torgersen  2007           39.5
##  3 Adelie  Torgersen  2007           40.3
##  4 Adelie  Torgersen  2007           NA
##  5 Adelie  Torgersen  2007           36.7
##  6 Adelie  Torgersen  2007           39.3
##  7 Adelie  Torgersen  2007           38.9
##  8 Adelie  Torgersen  2007           39.2
##  9 Adelie  Torgersen  2007           34.1
## 10 Adelie  Torgersen  2007           42
## # … with 334 more rows
```

```
select(penguins, species, island, year, starts_with("bill"))
```

```
## # A tibble: 344 x 5
##    species island     year bill_length_mm bill_depth_mm
##    <fct>   <fct>     <int>          <dbl>         <dbl>
##  1 Adelie  Torgersen  2007           39.1          18.7
##  2 Adelie  Torgersen  2007           39.5          17.4
##  3 Adelie  Torgersen  2007           40.3          18
##  4 Adelie  Torgersen  2007           NA            NA
##  5 Adelie  Torgersen  2007           36.7          19.3
##  6 Adelie  Torgersen  2007           39.3          20.6
##  7 Adelie  Torgersen  2007           38.9          17.8
##  8 Adelie  Torgersen  2007           39.2          19.6
##  9 Adelie  Torgersen  2007           34.1          18.1
## 10 Adelie  Torgersen  2007           42            20.2
## # … with 334 more rows
```

```
select(penguins, species, island, body_mass_g:year, ends_with("mm"))
```

```
## # A tibble: 344 x 8
##    species island body_mass_g sex    year bill_length_mm bill_depth_mm
##    <fct>   <fct>        <int> <fct> <int>          <dbl>         <dbl>
##  1 Adelie  Torge…        3750 male   2007           39.1          18.7
##  2 Adelie  Torge…        3800 fema…  2007           39.5          17.4
##  3 Adelie  Torge…        3250 fema…  2007           40.3          18
##  4 Adelie  Torge…          NA <NA>   2007           NA            NA
##  5 Adelie  Torge…        3450 fema…  2007           36.7          19.3
##  6 Adelie  Torge…        3650 male   2007           39.3          20.6
##  7 Adelie  Torge…        3625 fema…  2007           38.9          17.8
##  8 Adelie  Torge…        4675 male   2007           39.2          19.6
##  9 Adelie  Torge…        3475 <NA>   2007           34.1          18.1
## 10 Adelie  Torge…        4250 <NA>   2007           42            20.2
## # … with 334 more rows, and 1 more variable: flipper_length_mm <int>
```

```
select(penguins, species, ends_with("mm"))
```

```
## # A tibble: 344 x 4
##    species bill_length_mm bill_depth_mm flipper_length_mm
##    <fct>            <dbl>         <dbl>             <int>
##  1 Adelie            39.1          18.7               181
##  2 Adelie            39.5          17.4               186
##  3 Adelie            40.3          18                 195
##  4 Adelie            NA            NA                  NA
##  5 Adelie            36.7          19.3               193
##  6 Adelie            39.3          20.6               190
##  7 Adelie            38.9          17.8               181
##  8 Adelie            39.2          19.6               195
##  9 Adelie            34.1          18.1               193
## 10 Adelie            42            20.2               190
## # … with 334 more rows
```

```
select(penguins, is.numeric)
```

```
## # A tibble: 344 x 5
##    bill_length_mm bill_depth_mm flipper_length_mm body_mass_g  year
##             <dbl>         <dbl>             <int>       <int> <int>
##  1           39.1          18.7               181        3750  2007
##  2           39.5          17.4               186        3800  2007
##  3           40.3          18                 195        3250  2007
##  4           NA            NA                  NA          NA  2007
##  5           36.7          19.3               193        3450  2007
##  6           39.3          20.6               190        3650  2007
##  7           38.9          17.8               181        3625  2007
##  8           39.2          19.6               195        4675  2007
##  9           34.1          18.1               193        3475  2007
## 10           42            20.2               190        4250  2007
## # … with 334 more rows
```

```
select(penguins, !is.numeric)
```

```
## # A tibble: 344 x 3
##    species island    sex
##    <fct>   <fct>     <fct>
##  1 Adelie  Torgersen male
##  2 Adelie  Torgersen female
##  3 Adelie  Torgersen female
##  4 Adelie  Torgersen <NA>
##  5 Adelie  Torgersen female
##  6 Adelie  Torgersen male
##  7 Adelie  Torgersen female
##  8 Adelie  Torgersen male
##  9 Adelie  Torgersen <NA>
## 10 Adelie  Torgersen <NA>
## # … with 334 more rows
```

pipe

```
penguins
```

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

```
mutate(penguins, extra_col = 1)
```

same as:

```
penguins %>%
  mutate(extra_col = 1)
```

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

```r
penguins %>%
  select(1:2, contains("bill")) %>%
  mutate(extra_col = 1)
```

```
## # A tibble: 344 x 5
##    species island    bill_length_mm bill_depth_mm extra_col
##    <fct>   <fct>              <dbl>         <dbl>     <dbl>
##  1 Adelie  Torgersen           39.1          18.7         1
##  2 Adelie  Torgersen           39.5          17.4         1
##  3 Adelie  Torgersen           40.3          18           1
##  4 Adelie  Torgersen           NA            NA           1
##  5 Adelie  Torgersen           36.7          19.3         1
##  6 Adelie  Torgersen           39.3          20.6         1
##  7 Adelie  Torgersen           38.9          17.8         1
##  8 Adelie  Torgersen           39.2          19.6         1
##  9 Adelie  Torgersen           34.1          18.1         1
## 10 Adelie  Torgersen           42            20.2         1
## # … with 334 more rows
```

```
penguins %>%
  select(1:2, contains("bill")) %>%
  mutate(extra_col = 1) %>%
  filter(sex == "female")
```

mutate

```
penguins %>%
  mutate(
    new_column = 1,
    bill_ld_ratio = bill_length_mm/bill_depth_mm
    )
```

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

```
penguins %>%
  group_by(species) %>%
  mutate(
    bill_length_sp_max = max(bill_length_mm, na.rm = TRUE),
    bill_length_pc = (bill_length_mm/bill_length_sp_max)*100
    )
```

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

```
penguins %>%
  mutate(across(ends_with("mm"),
                ~ .x/10))
```

```
## # A tibble: 344 x 8
##    species island bill_length_mm bill_depth_mm flipper_length_… body_mass_g
##    <fct>   <fct>           <dbl>         <dbl>             <dbl>       <int>
##  1 Adelie  Torge…           3.91          1.87              18.1        3750
##  2 Adelie  Torge…           3.95          1.74              18.6        3800
##  3 Adelie  Torge…           4.03          1.8               19.5        3250
##  4 Adelie  Torge…          NA            NA                 NA            NA
##  5 Adelie  Torge…           3.67          1.93              19.3        3450
##  6 Adelie  Torge…           3.93          2.06              19          3650
##  7 Adelie  Torge…           3.89          1.78              18.1        3625
##  8 Adelie  Torge…           3.92          1.96              19.5        4675
##  9 Adelie  Torge…           3.41          1.81              19.3        3475
## 10 Adelie  Torge…           4.2           2.02              19          4250
## # … with 334 more rows, and 2 more variables: sex <fct>, year <int>
```