

Data Viz Grundlagen

Dr. Shirin Elsinghorst

4/8/2021

Libraries

```
library(tidyverse)

library(ggExtra)
library(ragg)
library(ggalluvial)
library(treemapify)
library(ggalt)

library(palmerpenguins)
```

Dataset

```
head(penguins)

## # A tibble: 6 x 8
##   species island bill_length_mm bill_depth_mm flipper_length_~ body_mass_g sex
##   <fct>   <fct>         <dbl>         <dbl>         <int>         <int> <fct>
## 1 Adelie  Torge~           39.1           18.7           181           3750 male
## 2 Adelie  Torge~           39.5           17.4           186           3800 fema~
## 3 Adelie  Torge~           40.3            18           195           3250 fema~
## 4 Adelie  Torge~            NA            NA            NA            NA <NA>
## 5 Adelie  Torge~           36.7           19.3           193           3450 fema~
## 6 Adelie  Torge~           39.3           20.6           190           3650 male
## # ... with 1 more variable: year <int>
```

```
#head(penguins_raw)
```

Colors

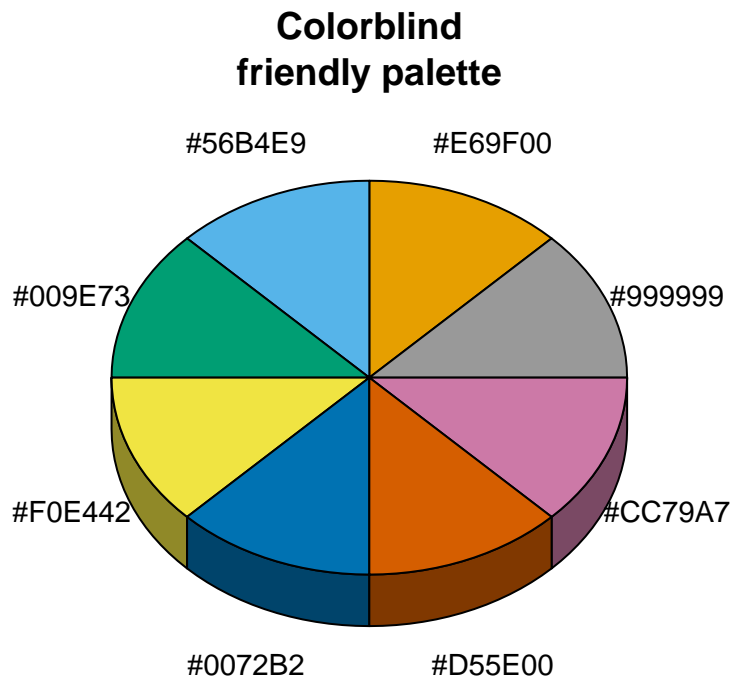
- set colorblind-friendly palettes

```
# The palette with grey:
cbp1 <- c("#999999", "#E69F00", "#56B4E9", "#009E73",
          "#F0E442", "#0072B2", "#D55E00", "#CC79A7")

# The palette with black:
cbp2 <- c("#000000", "#E69F00", "#56B4E9", "#009E73",
          "#F0E442", "#0072B2", "#D55E00", "#CC79A7")

library(plotrix)
sliceValues <- rep(10, 8) # each slice value=10 for proportionate slices
```

```
(
  p <- pie3D(sliceValues,
    explode=0,
    theta = 1.2,
    col = cbp1,
    labels = cbp1,
    labelcex = 0.9,
    shade = 0.6,
    main = "Colorblind\nfriendly palette")
)
```



```
## [1] 0.3926991 1.1780972 1.9634954 2.7488936 3.5342917 4.3196899 5.1050881
## [8] 5.8904862
```

```
ggplot <- function(...) ggplot2::ggplot(...) +
  scale_color_manual(values = cbp1) +
  scale_fill_manual(values = cbp1) + # note: needs to be overridden when using continuous color scales
  theme_bw()
```

Tidyverse

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>
```

```
penguins %>%
  gather("key", "value", bill_length_mm:body_mass_g) %>%
  head()
```

```
## # A tibble: 6 x 6
##   species island    sex    year key          value
##   <fct>   <fct>   <fct> <int> <chr>        <dbl>
## 1 Adelie Torgersen male    2007 bill_length_mm 39.1
## 2 Adelie Torgersen female 2007 bill_length_mm 39.5
## 3 Adelie Torgersen female 2007 bill_length_mm 40.3
## 4 Adelie Torgersen <NA>    2007 bill_length_mm NA
## 5 Adelie Torgersen female 2007 bill_length_mm 36.7
## 6 Adelie Torgersen male    2007 bill_length_mm 39.3
```

```
penguins %>%
  group_by(species, sex) %>%
  count(island) %>%
  arrange(desc(n))
```

```
## # A tibble: 13 x 4
## # Groups:   species, sex [8]
##   species sex    island    n
##   <fct>   <fct> <fct>   <int>
## 1 Gentoo male    Biscoe    61
## 2 Gentoo female Biscoe    58
## 3 Chinstrap female Dream     34
## 4 Chinstrap male  Dream     34
## 5 Adelie male    Dream     28
## 6 Adelie female Dream     27
## 7 Adelie female Torgersen 24
## 8 Adelie male    Torgersen 23
## 9 Adelie female Biscoe    22
## 10 Adelie male    Biscoe    22
## 11 Adelie <NA>    Torgersen 5
## 12 Gentoo <NA>    Biscoe    5
## 13 Adelie <NA>    Dream     1
```

```
penguins %>%
  gather("key", "value", bill_length_mm:body_mass_g) %>%
  group_by(species, sex, island, key) %>%
  summarise(n = n(),
            sum = sum(value),
            mean = mean(value),
            median = median(value),
            sd = sd(value),
            se = sd(value) / sqrt(n()))
```

```
## # A tibble: 52 x 10
## # Groups:   species, sex, island [13]
##   species sex    island key          n    sum  mean median    sd    se
```

```
##      <fct>   <fct> <fct>      <chr>      <int>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
##  1 Adelie   female Biscoe  bill_depth~  22   390.   17.7   17.7   1.09   0.233
##  2 Adelie   female Biscoe  bill_lengt~  22   822.   37.4   37.8   1.76   0.376
##  3 Adelie   female Biscoe  body_mass_g  22  74125  3369.  3375   343.   73.2
##  4 Adelie   female Biscoe  flipper_le~  22   4118   187.   187    6.74   1.44
##  5 Adelie   female Dream  bill_depth~  27   476.   17.6   17.8   0.897   0.173
##  6 Adelie   female Dream  bill_lengt~  27   997.   36.9   36.8   2.09   0.402
##  7 Adelie   female Dream  body_mass_g  27  90300  3344.  3400   212.   40.8
##  8 Adelie   female Dream  flipper_le~  27   5072   188.   188    5.51   1.06
##  9 Adelie   female Torgers~ bill_depth~  24   421.   17.6   17.4   0.880   0.180
## 10 Adelie   female Torgers~ bill_lengt~  24   901.   37.6   37.6   2.21   0.451
## # ... with 42 more rows
```

```
penguins %>%
  group_by(species, sex, island, year) %>%
  summarise_each(funs(sum,
                        mean,
                        median,
                        sd,
                        se = sd(.) / sqrt(n())
                      ))
```

```
## # A tibble: 35 x 24
## # Groups:   species, sex, island [13]
##   species sex  island  year bill_length_mm~ bill_depth_mm_s~ flipper_length~
##   <fct>   <fct> <fct>  <int>      <dbl>          <dbl>          <int>
##  1 Adelie fema~ Biscoe  2007      187.           92.9           909
##  2 Adelie fema~ Biscoe  2008      330.           155           1679
##  3 Adelie fema~ Biscoe  2009      305.           142.           1530
##  4 Adelie fema~ Dream   2007      341.           161.           1665
##  5 Adelie fema~ Dream   2008      290.           142.           1512
##  6 Adelie fema~ Dream   2009      366.           173.           1895
##  7 Adelie fema~ Torge~  2007      306.           145.           1501
##  8 Adelie fema~ Torge~  2008      293.           139.           1520
##  9 Adelie fema~ Torge~  2009      302.           137.           1498
## 10 Adelie male~ Biscoe  2007      196.           91.5           908
## # ... with 25 more rows, and 17 more variables: body_mass_g_sum <int>,
## #   bill_length_mm_mean <dbl>, bill_depth_mm_mean <dbl>,
## #   flipper_length_mm_mean <dbl>, body_mass_g_mean <dbl>,
## #   bill_length_mm_median <dbl>, bill_depth_mm_median <dbl>,
## #   flipper_length_mm_median <dbl>, body_mass_g_median <dbl>,
## #   bill_length_mm_sd <dbl>, bill_depth_mm_sd <dbl>,
## #   flipper_length_mm_sd <dbl>, body_mass_g_sd <dbl>, bill_length_mm_se <dbl>,
## #   bill_depth_mm_se <dbl>, flipper_length_mm_se <dbl>, body_mass_g_se <dbl>
```

Main diagram types

Pointcharts

```
penguins %>%
  remove_missing() %>%
  ggplot(aes(x = bill_length_mm, y = flipper_length_mm)) +
  geom_jitter(alpha = 0.5) +
  facet_wrap(vars(species), ncol = 3) +
```

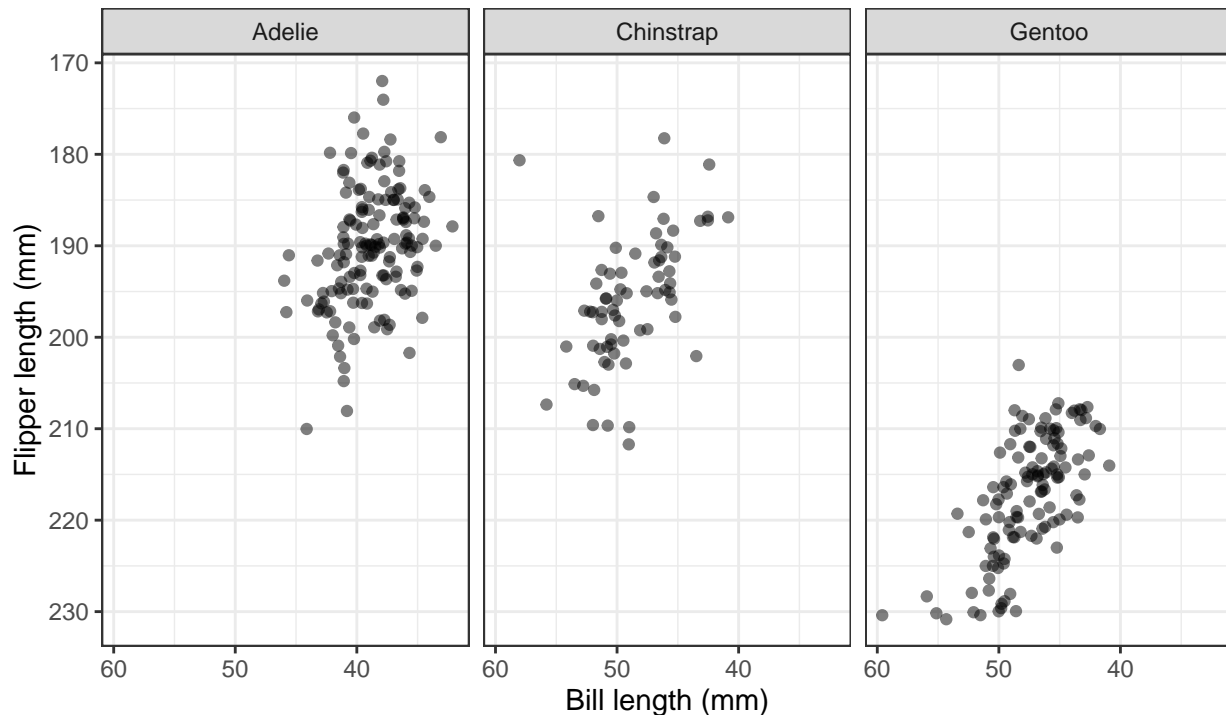
```

scale_x_reverse() +
scale_y_reverse() +
labs(x = "Bill length (mm)",
     y = "Flipper length (mm)",
     size = "body mass (g)",
     title = "Scatterplot",
     subtitle = "Penguins bill v. flipper length by species",
     caption = "Source: https://github.com/allisonhorst/palmerpenguins")

```

Scatterplot

Penguins bill v. flipper length by species



Source: <https://github.com/allisonhorst/palmerpenguins>

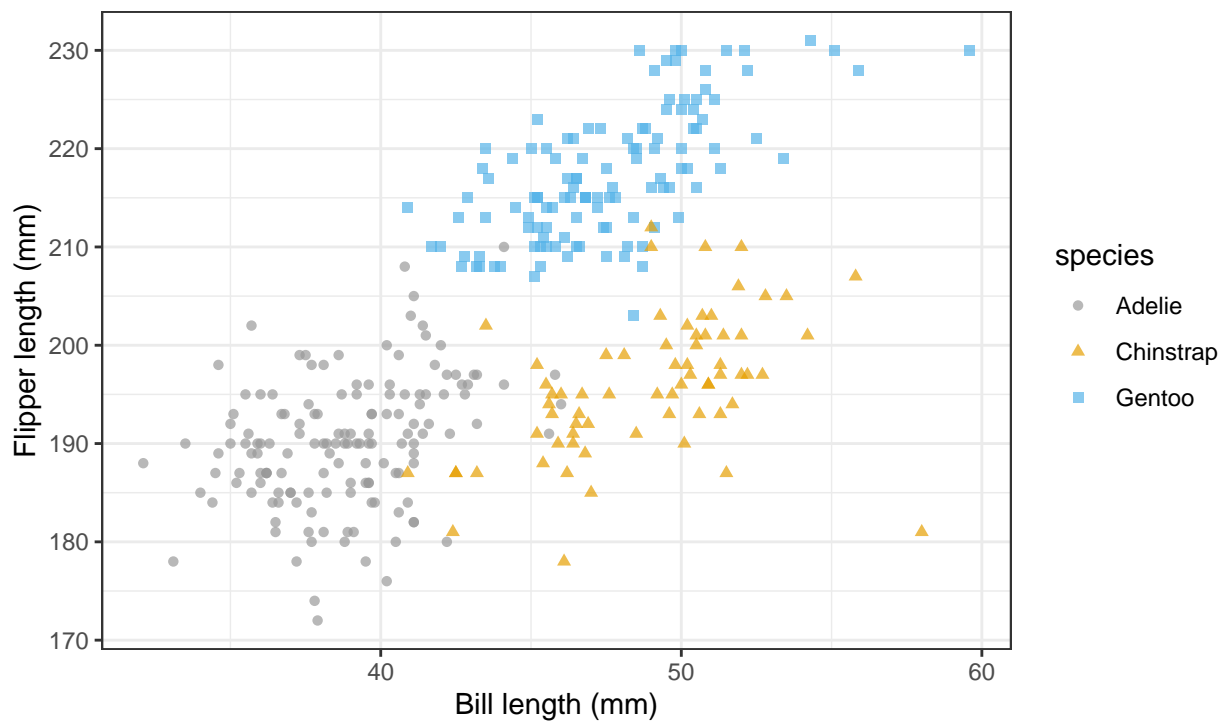
```

penguins %>%
  remove_missing() %>%
  ggplot(aes(x = bill_length_mm, y = flipper_length_mm,
             color = species, shape = species)) +
  geom_point(alpha = 0.7) +
  labs(x = "Bill length (mm)",
       y = "Flipper length (mm)",
       title = "Scatterplot",
       subtitle = "Penguins bill v. flipper length by species",
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")

```

Scatterplot

Penguins bill v. flipper length by species



Source: <https://github.com/allisonhorst/palmerpenguins>

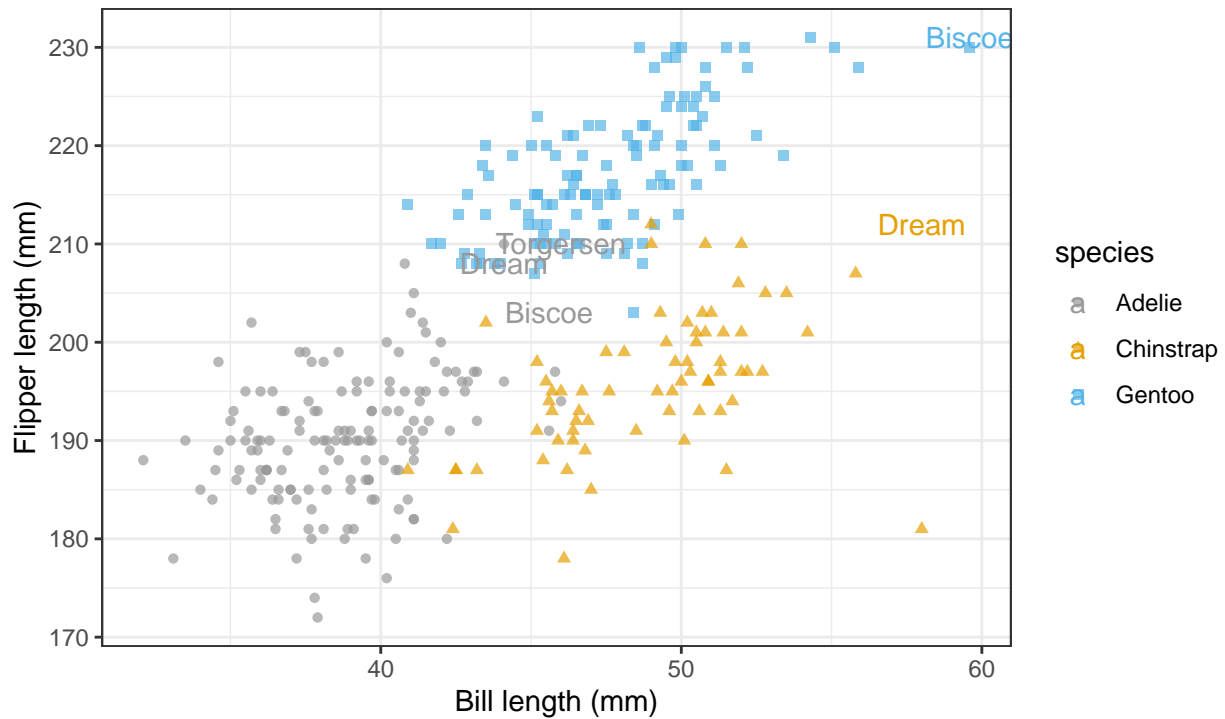
- with labels/text

```
max_labels <- penguins %>%  
  remove_missing() %>%  
  group_by(species, island) %>%  
  summarise(bill_length_mm = max(bill_length_mm),  
            flipper_length_mm = max(flipper_length_mm))
```

```
penguins %>%  
  remove_missing() %>%  
  ggplot(aes(x = bill_length_mm, y = flipper_length_mm,  
            color = species, shape = species)) +  
  geom_point(alpha = 0.7) +  
  geom_text(data = max_labels, aes(label = island)) +  
  labs(x = "Bill length (mm)",  
       y = "Flipper length (mm)",  
       title = "Scatterplot",  
       subtitle = "Penguins bill v. flipper length by species",  
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```

Scatterplot

Penguins bill v. flipper length by species



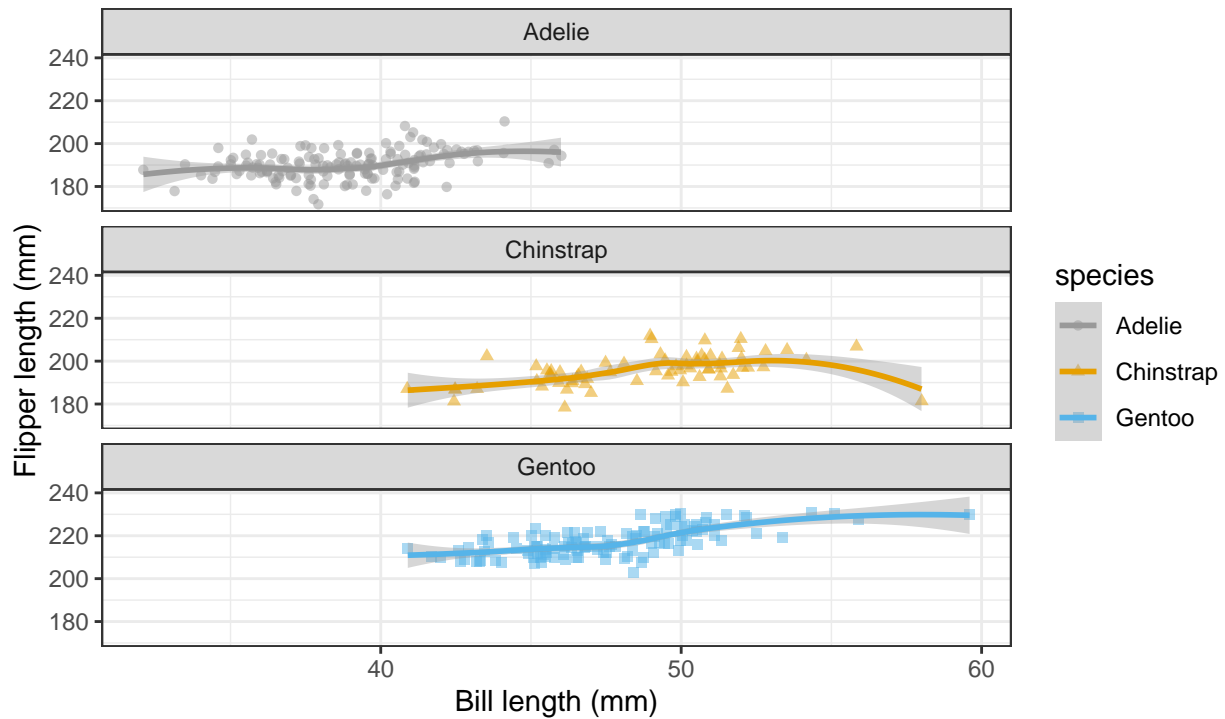
Source: <https://github.com/allisonhorst/palmerpenguins>

- Jitter with smoothing line

```
penguins %>%
  remove_missing() %>%
  ggplot(aes(x = bill_length_mm, y = flipper_length_mm,
             color = species, shape = species)) +
  geom_jitter(alpha = 0.5) +
  geom_smooth(method = "loess", se = TRUE) +
  facet_wrap(vars(species), nrow = 3) +
  labs(x = "Bill length (mm)",
       y = "Flipper length (mm)",
       title = "Scatterplot with smoothing line",
       subtitle = "Penguins bill v. flipper length by species with loess smoothing line",
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```

Scatterplot with smoothing line

Penguins bill v. flipper length by species with loess smoothing line

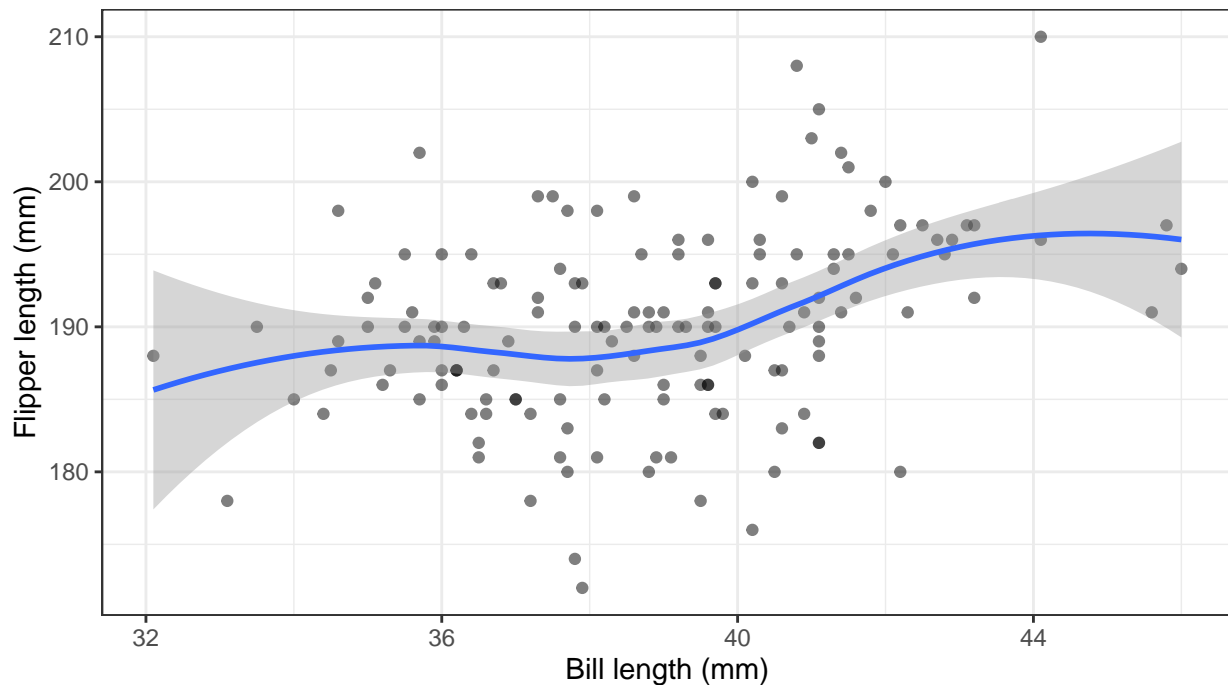


Source: <https://github.com/allisonhorst/palmerpenguins>

```
penguins %>%
  remove_missing() %>%
  filter(species == "Adelie") %>%
  ggplot(aes(x = bill_length_mm, y = flipper_length_mm)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "loess", se = TRUE) +
  labs(x = "Bill length (mm)",
       y = "Flipper length (mm)",
       title = "Scatterplot with smoothing line",
       subtitle = "Penguins bill v. flipper length by species with\nloess smoothing line, histogram & c",
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```


Scatterplot with smoothing line

Penguins bill v. flipper length by species with
loess smoothing line, histogram & density distribution



Source: <https://github.com/allisonhorst/palmerpenguins>

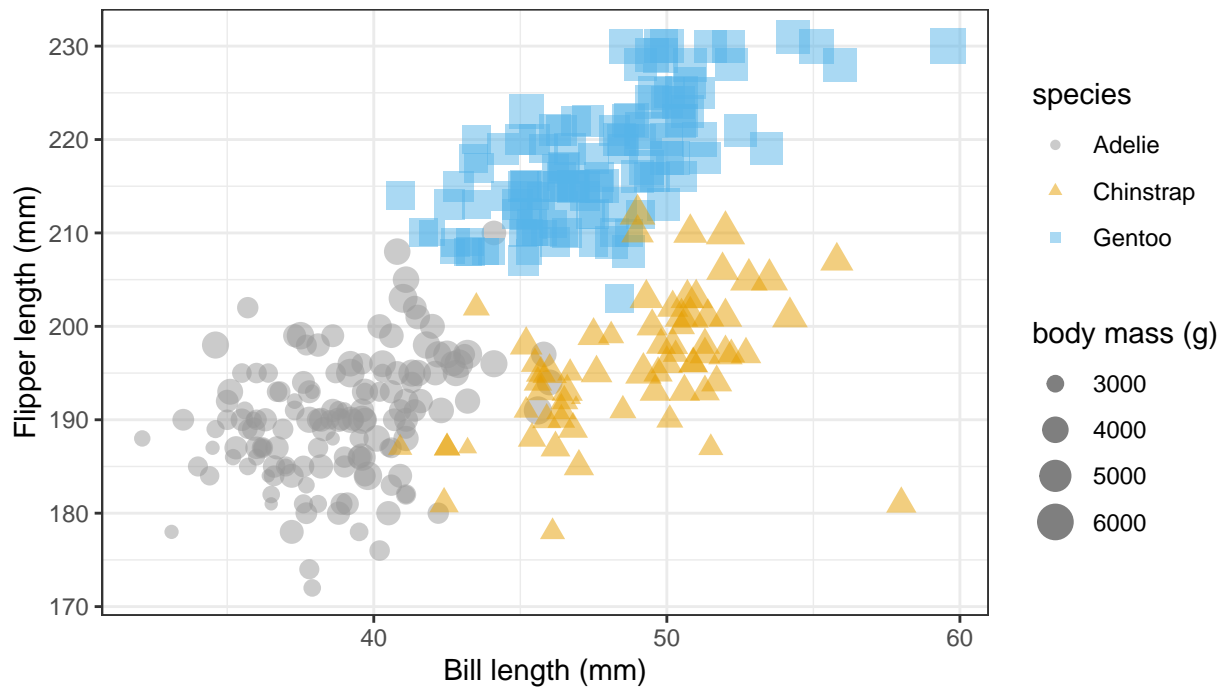
```
#(ggMarginal(p, type = "densigram", fill = "transparent"))
```

Bubblecharts

```
penguins %>%  
  remove_missing() %>%  
  ggplot(aes(x = bill_length_mm, y = flipper_length_mm,  
             color = species, shape = species, size = body_mass_g)) +  
  geom_point(alpha = 0.5) +  
  labs(x = "Bill length (mm)",  
       y = "Flipper length (mm)",  
       title = "Bubble plot",  
       size = "body mass (g)",  
       subtitle = "Penguins bill v. flipper length by species;\nsize indicates body mass in grams",  
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```

Bubble plot

Penguins bill v. flipper length by species;
size indicates body mass in grams



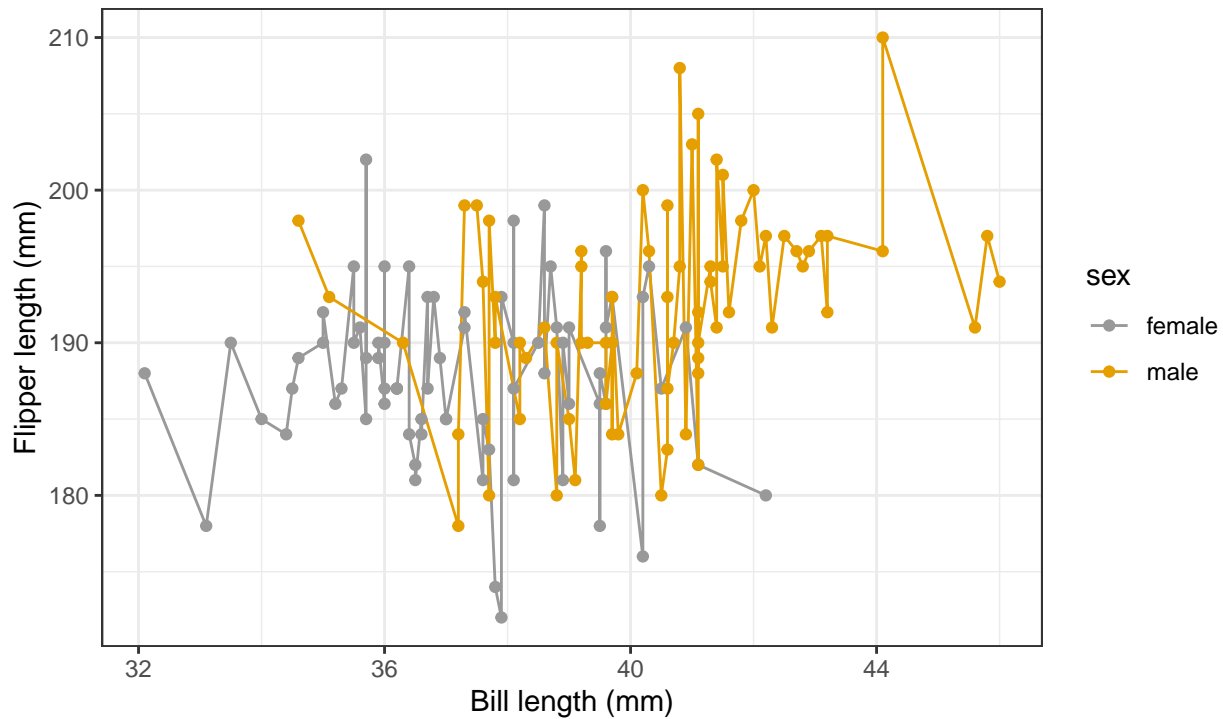
Source: <https://github.com/allisonhorst/palmerpenguins>

Linecharts

```
penguins %>%  
  remove_missing() %>%  
  filter(species == "Adelie") %>%  
  ggplot(aes(x = bill_length_mm, y = flipper_length_mm,  
             color = sex)) +  
  geom_line() +  
  geom_point() +  
  labs(x = "Bill length (mm)",  
       y = "Flipper length (mm)",  
       title = "Line plot",  
       subtitle = "Penguins bill v. flipper length by species and sex",  
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```

Line plot

Penguins bill v. flipper length by species and sex



Source: <https://github.com/allisonhorst/palmerpenguins>

Correlation plots / heatmaps

```
mat <- penguins %>%
  remove_missing() %>%
  select(bill_depth_mm, bill_length_mm, body_mass_g, flipper_length_mm)

cormat <- round(cor(mat), 2)
cormat[upper.tri(cormat)] <- NA

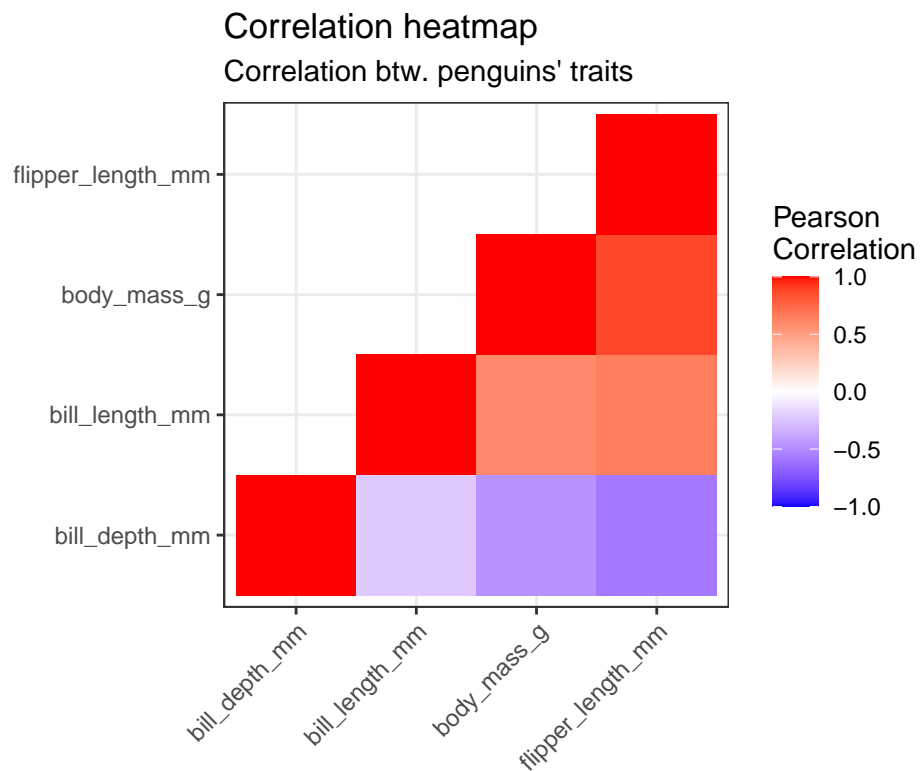
cormat <- cormat %>%
  as_data_frame() %>%
  mutate(x = colnames(mat)) %>%
  gather(key = "y", value = "value", bill_depth_mm:flipper_length_mm)

cormat %>%
  remove_missing() %>%
  arrange(x, y) %>%
  ggplot(aes(x = x, y = y, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(low = "blue", high = "red", mid = "white",
    midpoint = 0, limit = c(-1,1), space = "Lab",
    name = "Pearson\nCorrelation") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  coord_fixed() +
  labs(x = "",
    y = "",
```

```

title = "Correlation heatmap",
subtitle = "Correlation btw. penguins' traits",
caption = "Source: https://github.com/allisonhorst/palmerpenguins")

```



Source: <https://github.com/allisonhorst/palmerpenguins>

Barcharts

- per default: counts

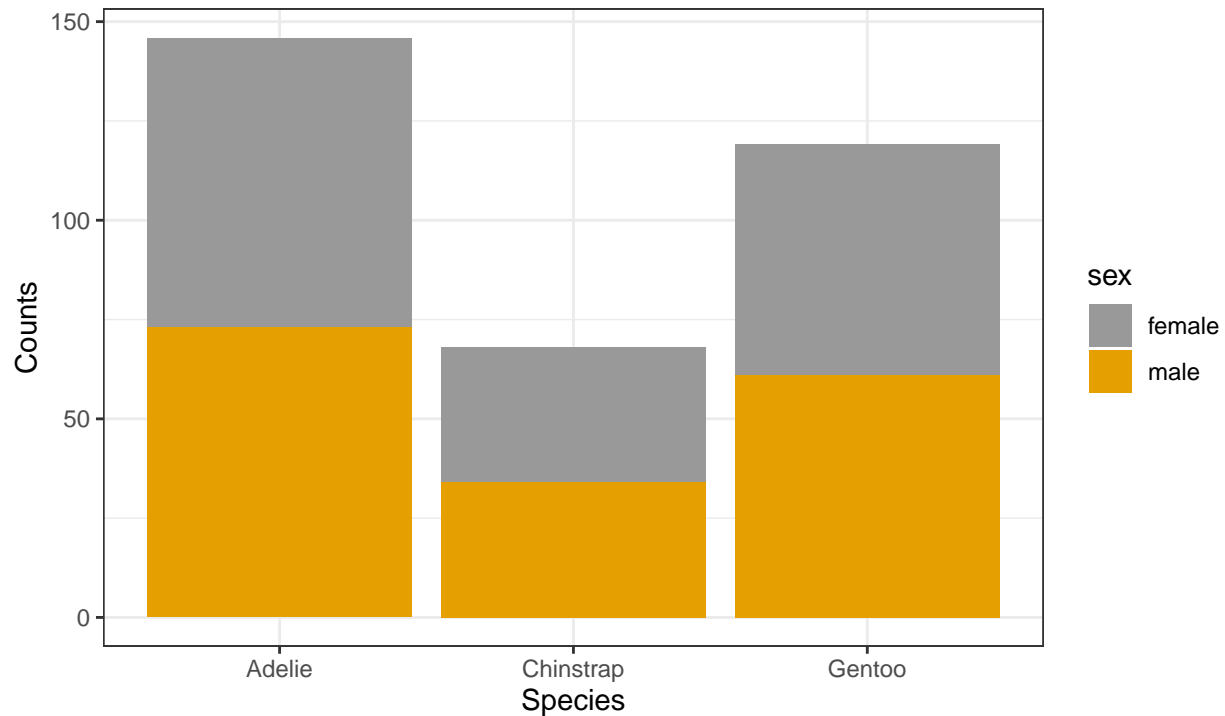
```

penguins %>%
  remove_missing() %>%
  ggplot(aes(x = species,
             fill = sex)) +
  geom_bar() +
  labs(x = "Species",
       y = "Counts",
       title = "Barchart",
       subtitle = "Counts of male & female penguins per species in study",
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")

```

Barchart

Counts of male & female penguins per species in study

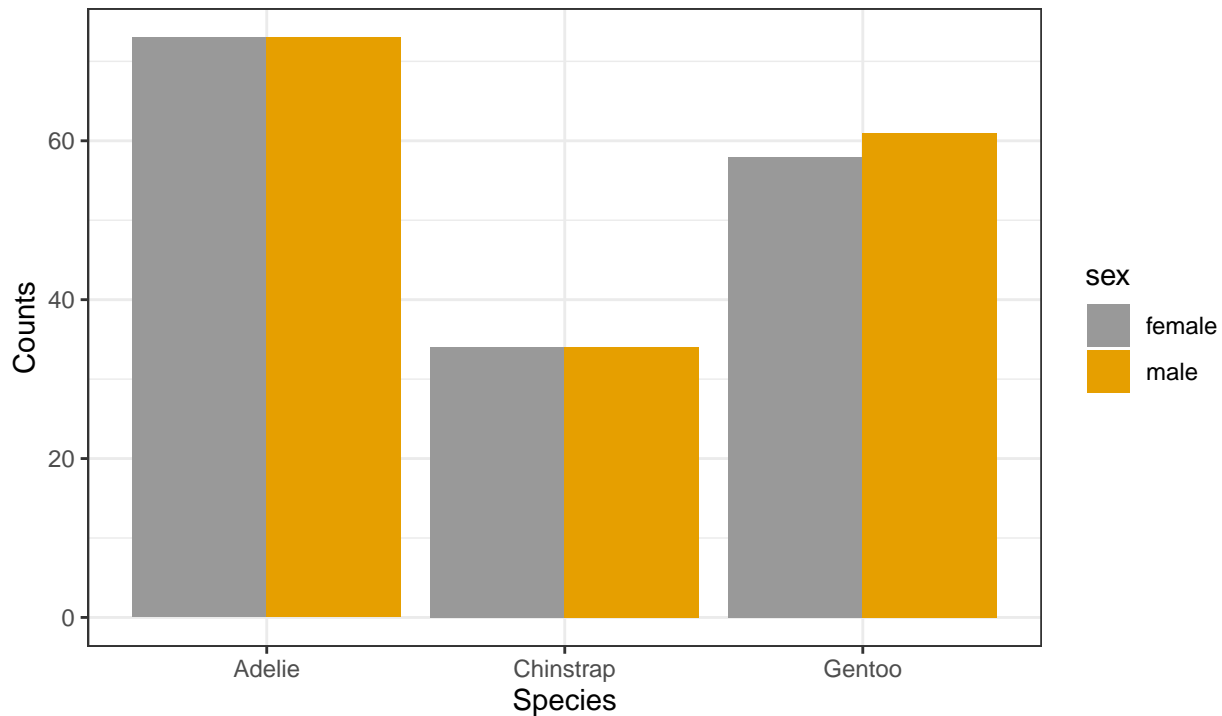


Source: <https://github.com/allisonhorst/palmerpenguins>

```
penguins %>%  
  remove_missing() %>%  
  ggplot(aes(x = species,  
             fill = sex)) +  
  geom_bar(position = 'dodge') +  
  labs(x = "Species",  
       y = "Counts",  
       title = "Barchart",  
       subtitle = "Counts of male & female penguins per species in study",  
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```

Barchart

Counts of male & female penguins per species in study



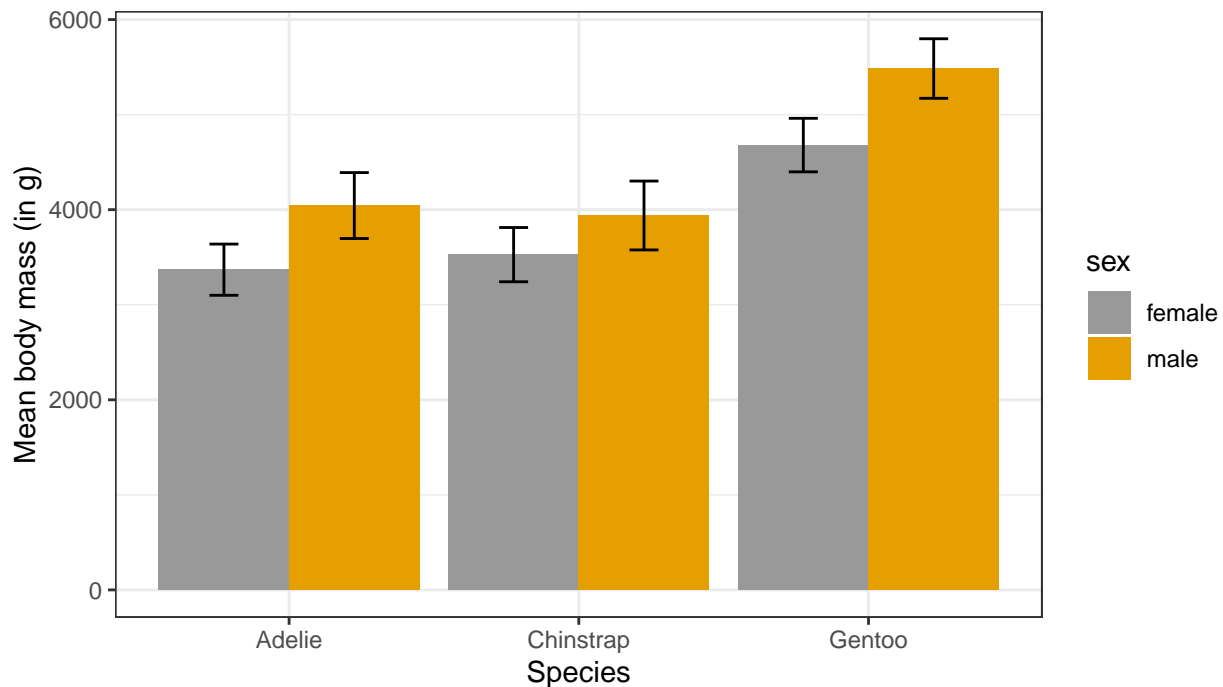
Source: <https://github.com/allisonhorst/palmerpenguins>

- alternative: set y-values

```
penguins %>%
  remove_missing() %>%
  group_by(species, sex) %>%
  summarise(mean_bmg = mean(body_mass_g),
            sd_bmg = sd(body_mass_g)) %>%
  ggplot(aes(x = species, y = mean_bmg,
            fill = sex)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(aes(ymin = mean_bmg - sd_bmg,
                    ymax = mean_bmg + sd_bmg),
                width = 0.2,
                position = position_dodge(0.9)) +
  labs(x = "Species",
       y = "Mean body mass (in g)",
       title = "Barchart",
       subtitle = "Mean body mass of male & female penguins per species\nwith standard deviation",
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```

Barchart

Mean body mass of male & female penguins per species
with standard deviation



Source: <https://github.com/allisonhorst/palmerpenguins>

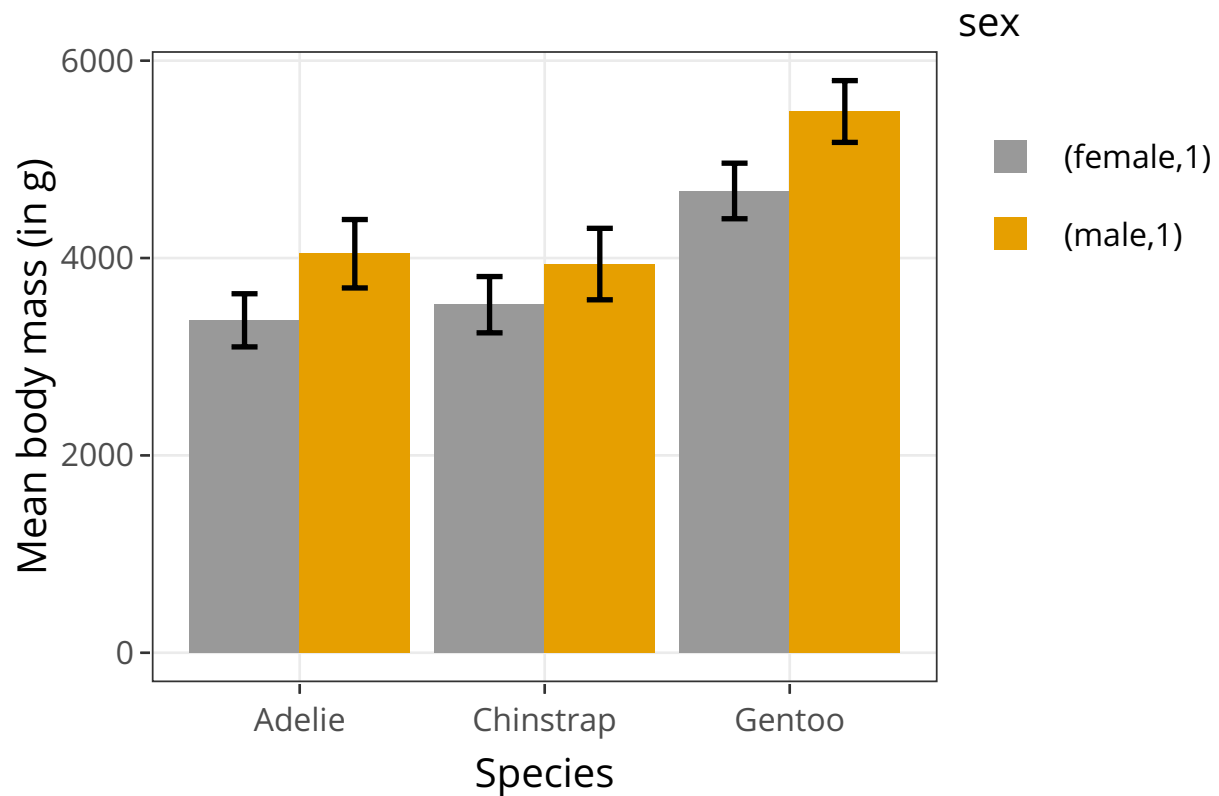
Plotly

```
library(plotly)

p <- penguins %>%
  remove_missing() %>%
  group_by(species, sex) %>%
  summarise(mean_bmg = mean(body_mass_g),
            sd_bmg = sd(body_mass_g)) %>%
  ggplot(aes(x = species, y = mean_bmg,
            fill = sex)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(aes(ymin = mean_bmg - sd_bmg,
                    ymax = mean_bmg + sd_bmg),
                width = 0.2,
                position = position_dodge(0.9)) +
  labs(x = "Species",
       y = "Mean body mass (in g)",
       title = "Barchart",
       subtitle = "Mean body mass of male & female penguins\nwith standard deviation",
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")

ggplotly(p)
```

Barchart

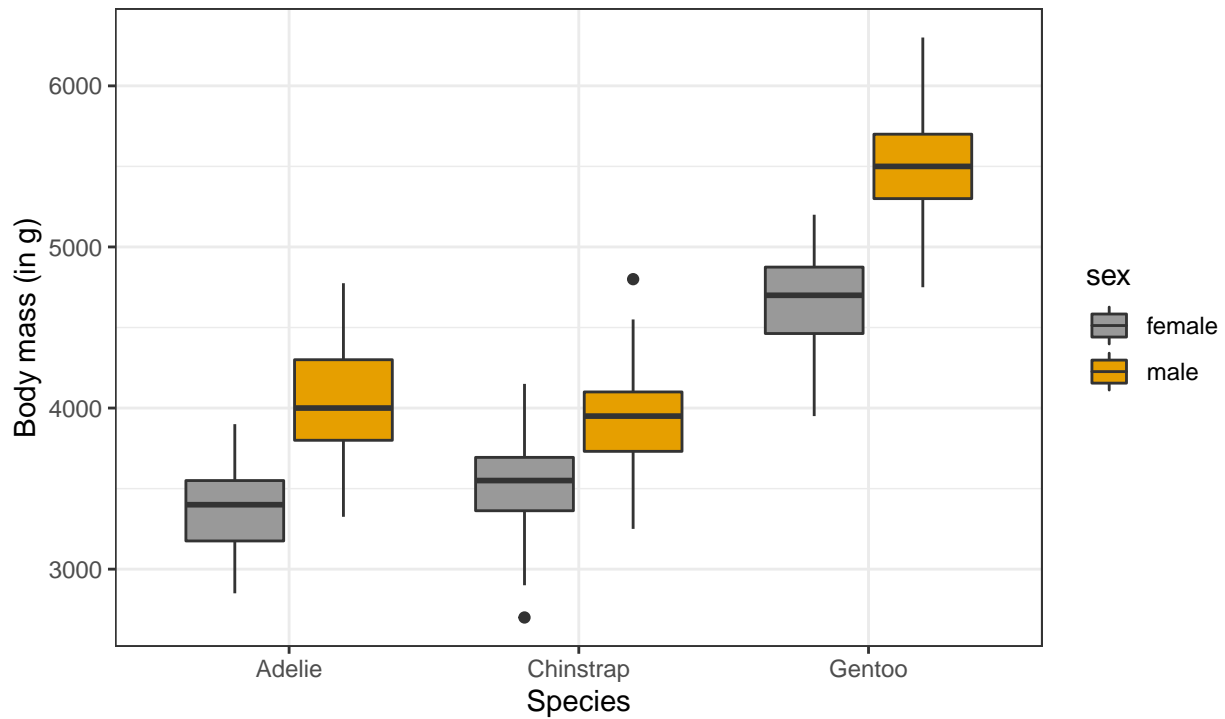


Boxplots

```
penguins %>%  
  remove_missing() %>%  
  ggplot(aes(x = species, y = body_mass_g,  
            fill = sex)) +  
  geom_boxplot() +  
  labs(x = "Species",  
       y = "Body mass (in g)",  
       title = "Boxplot",  
       subtitle = "Body mass of three penguin species per sex",  
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```


Boxplot

Body mass of three penguin species per sex



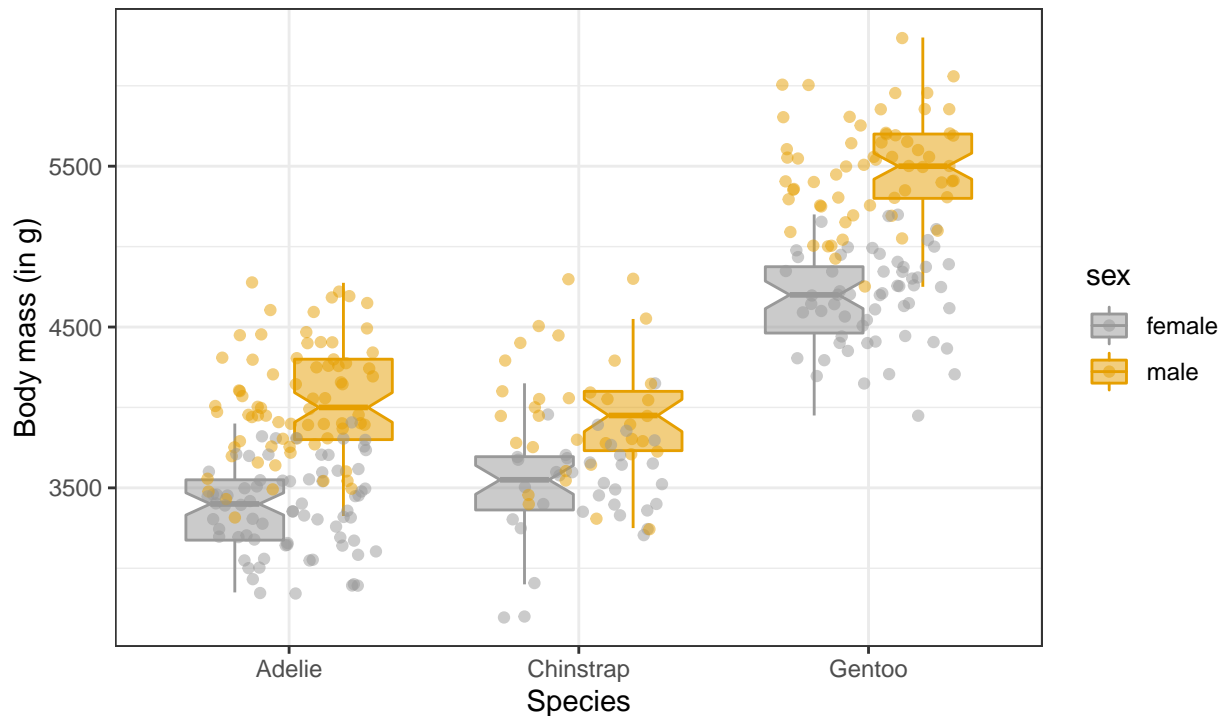
Source: <https://github.com/allisonhorst/palmerpenguins>

- with points

```
penguins %>%  
  remove_missing() %>%  
  ggplot(aes(x = species, y = body_mass_g,  
            fill = sex, color = sex)) +  
  geom_boxplot(alpha = 0.5, notch = TRUE) +  
  geom_jitter(alpha = 0.5, position=position_jitter(0.3)) +  
  labs(x = "Species",  
       y = "Body mass (in g)",  
       title = "Boxplot with points (dotplot)",  
       subtitle = "Body mass of three penguin species per sex",  
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```

Boxplot with points (dotplot)

Body mass of three penguin species per sex



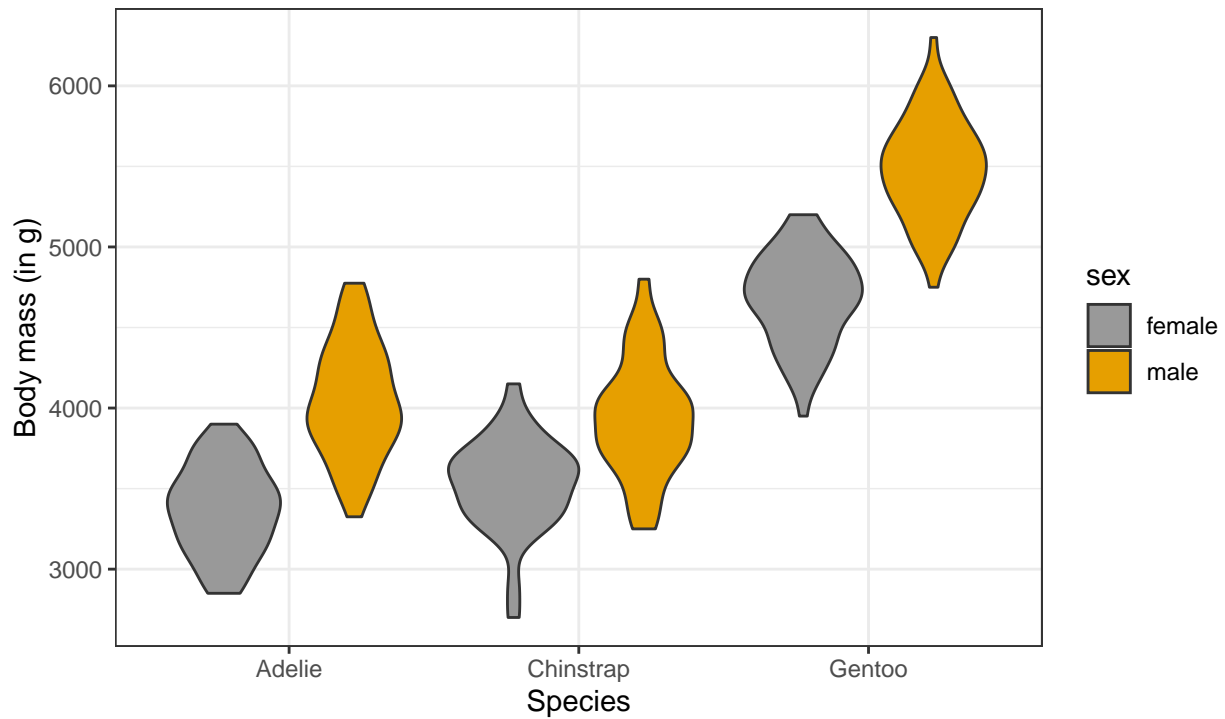
Source: <https://github.com/allisonhorst/palmerpenguins>

Violinplots

```
penguins %>%  
  remove_missing() %>%  
  ggplot(aes(x = species, y = body_mass_g,  
            fill = sex)) +  
  geom_violin(scale = "area") +  
  labs(x = "Species",  
       y = "Body mass (in g)",  
       title = "Violinplot",  
       subtitle = "Body mass of three penguin species per sex",  
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```

Violinplot

Body mass of three penguin species per sex



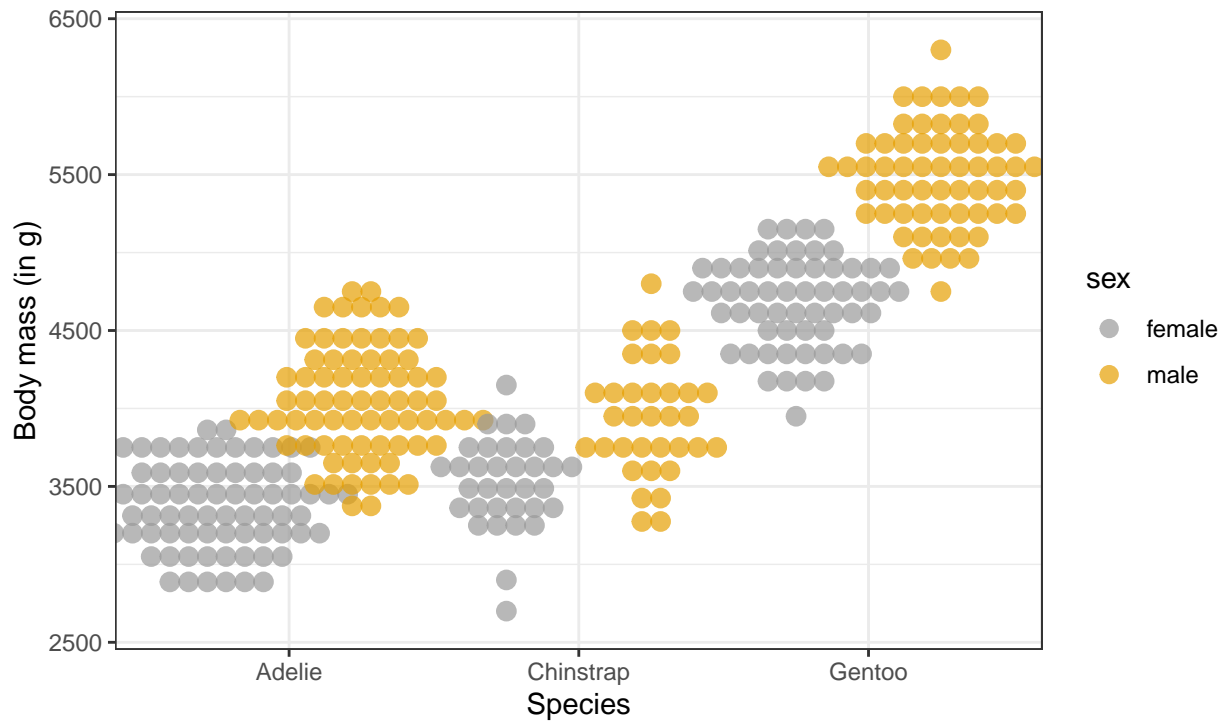
Source: <https://github.com/allisonhorst/palmerpenguins>

- with dots (sina-plots)

```
penguins %>%  
  remove_missing() %>%  
  ggplot(aes(x = species, y = body_mass_g,  
            fill = sex, color = sex)) +  
  geom_dotplot(method = "dotdensity", alpha = 0.7,  
              binaxis = 'y', stackdir = 'center',  
              position = position_dodge(1)) +  
  labs(x = "Species",  
       y = "Body mass (in g)",  
       title = "Violinplot with points (dotplot)",  
       subtitle = "Body mass of three penguin species per sex",  
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```

Violinplot with points (dotplot)

Body mass of three penguin species per sex



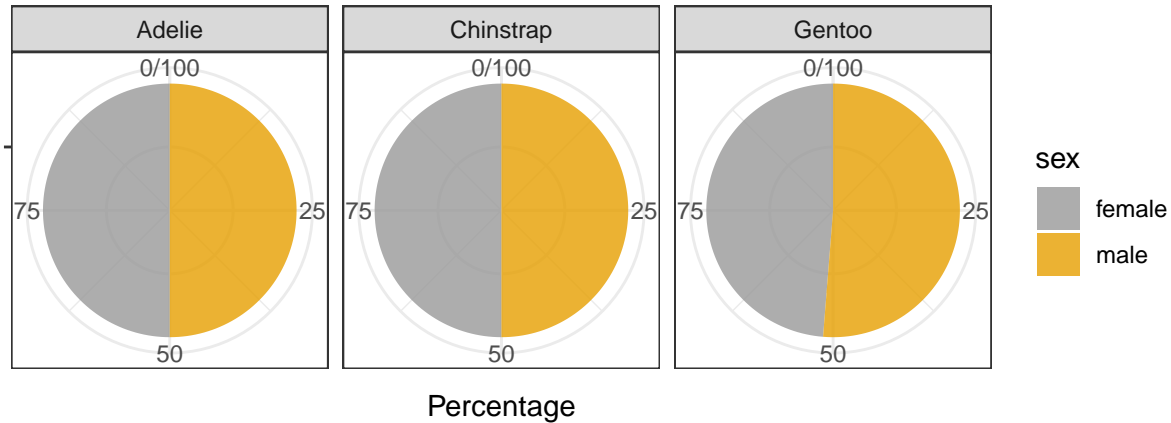
Source: <https://github.com/allisonhorst/palmerpenguins>

Piecharts

```
penguins %>%
  remove_missing() %>%
  group_by(species, sex) %>%
  summarise(n = n()) %>%
  mutate(freq = n / sum(n),
         percentage = freq * 100) %>%
  ggplot(aes(x = "", y = percentage,
            fill = sex)) +
  facet_wrap(vars(species), nrow = 1) +
  geom_bar(stat = "identity", alpha = 0.8) +
  coord_polar("y", start = 0) +
  labs(x = "",
       y = "Percentage",
       title = "Piechart",
       subtitle = "Percentage of male v. female penguins per species in study",
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```

Piechart

Percentage of male v. female penguins per species in study



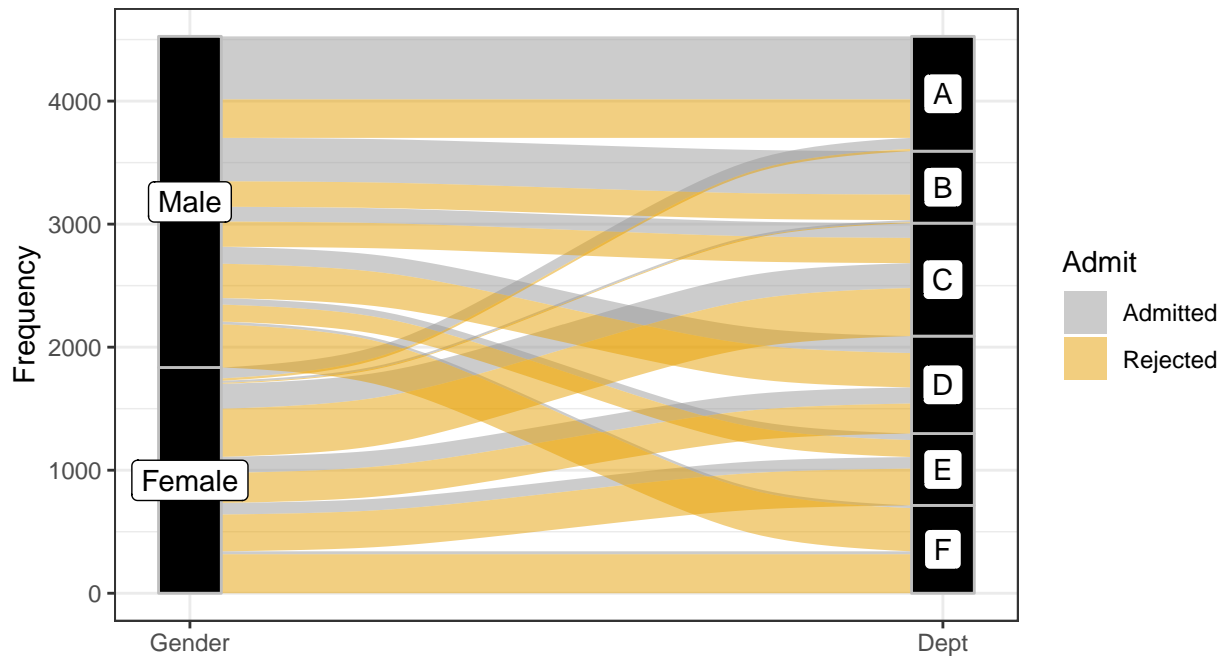
Source: <https://github.com/allisonhorst/palmerpenguins>

Alluvial charts

```
as.data.frame(UCBAdmissions) %>%
  ggplot(aes(y = Freq, axis1 = Gender, axis2 = Dept)) +
  geom_alluvium(aes(fill = Admit), width = 1/12) +
  geom_stratum(width = 1/12, fill = "black", color = "grey") +
  geom_label(stat = "stratum", aes(label = after_stat(stratum))) +
  scale_x_discrete(limits = c("Gender", "Dept"), expand = c(.05, .05)) +
  labs(x = "",
       y = "Frequency",
       title = "Alluvial chart",
       subtitle = "UC Berkeley admissions and rejections, by sex and department",
       caption = "Source: Bickel et al. (1975)\nSex bias in graduate admissions: Data from Berkeley.
```

Alluvial chart

UC Berkeley admissions and rejections, by sex and department



Source: Bickel et al. (1975)

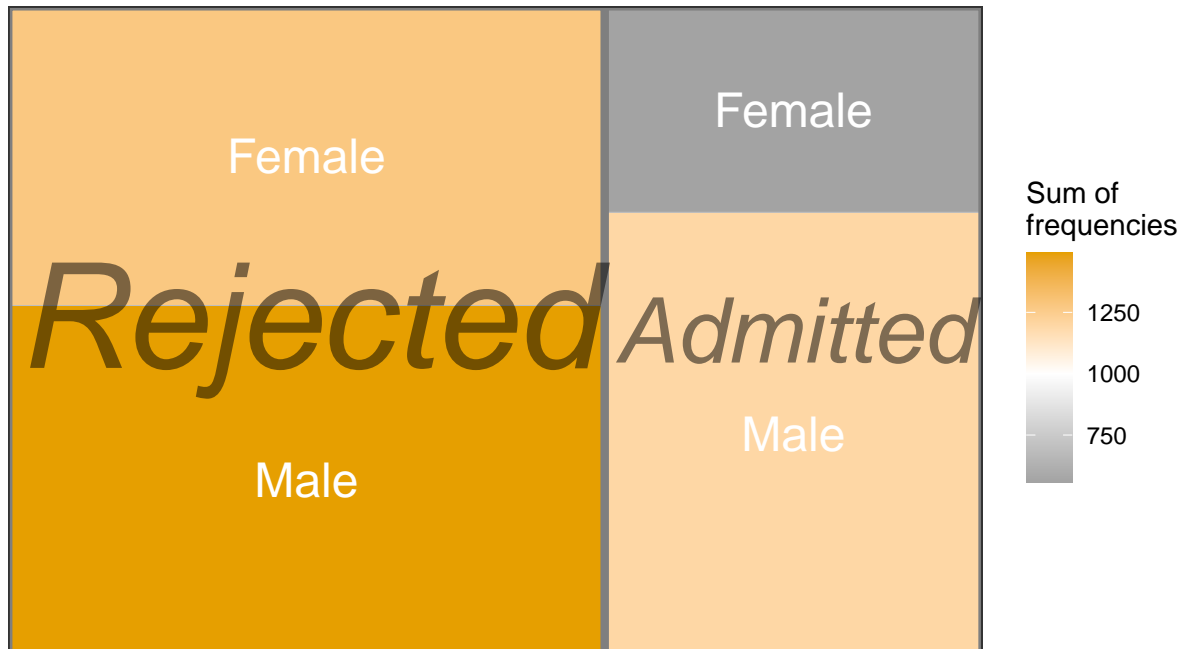
Sex bias in graduate admissions: Data from Berkeley. Science, 187, 398...403.

Treemaps

```
as.data.frame(UCBAdmissions) %>%
  group_by(Admit, Gender) %>%
  summarise(sum_freq = sum(Freq)) %>%
  ggplot(aes(area = sum_freq, fill = sum_freq, label = Gender,
             subgroup = Admit)) +
  geom_treemap() +
  geom_treemap_subgroup_border() +
  geom_treemap_subgroup_text(place = "centre", grow = T, alpha = 0.5, colour =
                             "black", fontface = "italic", min.size = 0) +
  geom_treemap_text(colour = "white", place = "centre", reflow = T) +
  scale_fill_gradient2(low = "#999999", high = "#E69F00", mid = "white", midpoint = 1000, space = "Lab") +
  name = "Sum of\nfrequencies") +
  labs(x = "",
       y = "",
       title = "Treemap",
       subtitle = "UC Berkeley admissions and rejections by sex",
       caption = "Source: Bickel et al. (1975)\nSex bias in graduate admissions: Data from Berkeley.
```

Treemap

UC Berkeley admissions and rejections by sex



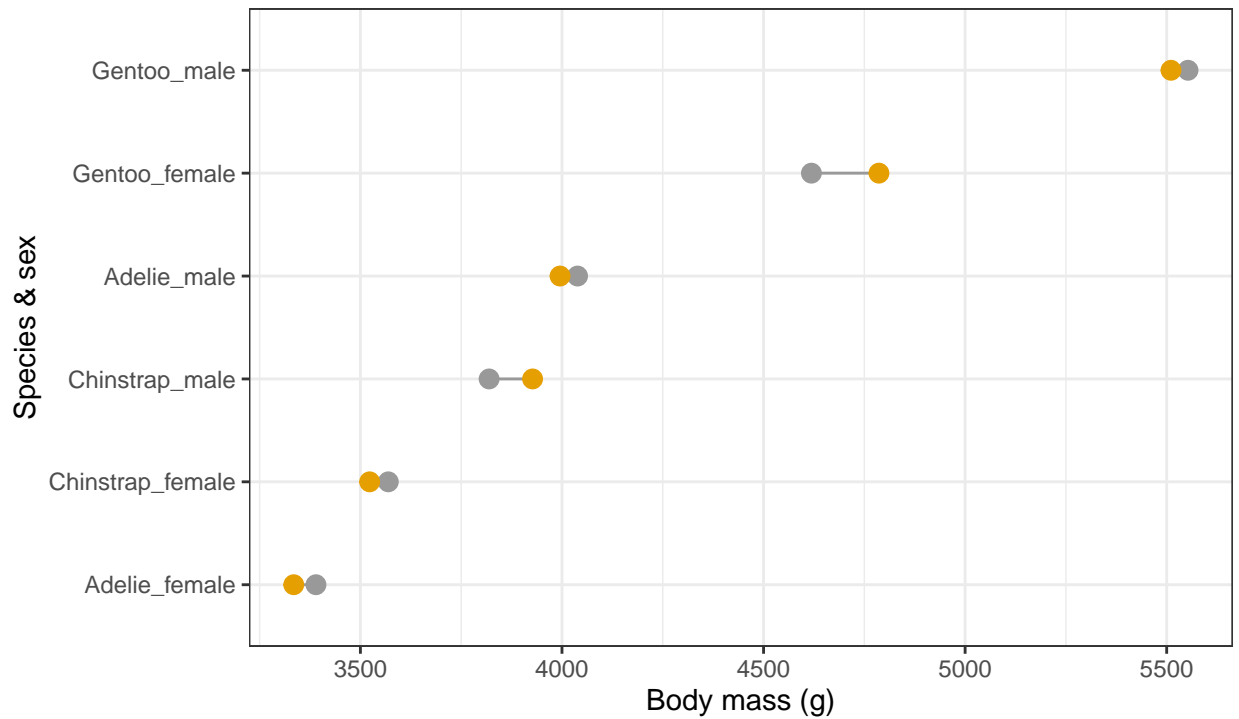
Source: Bickel et al. (1975)
Sex bias in graduate admissions: Data from Berkeley. Science, 187, 398...403.

Dumbbell plots

```
penguins %>%  
  remove_missing() %>%  
  group_by(year, species, sex) %>%  
  summarise(mean_bmg = mean(body_mass_g)) %>%  
  mutate(species_sex = paste(species, sex, sep = "_"),  
         year = paste0("year_", year)) %>%  
  spread(year, mean_bmg) %>%  
  ggplot(aes(x = year_2007, xend = year_2009,  
            y = reorder(species_sex, year_2009))) +  
  geom_dumbbell(color = "#999999",  
               size_x = 3,  
               size_xend = 3,  
               #Note: there is no US: 'color' for UK: 'colour'  
               # in geom_dumbbell unlike standard geoms in ggplot()  
               colour_x = "#999999",  
               colour_xend = "#E69F00") +  
  labs(x = "Body mass (g)",  
       y = "Species & sex",  
       title = "Dumbbell plot",  
       subtitle = "Penguin's change in body mass from 2007 to 2009",  
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```

Dumbbell plot

Penguin's change in body mass from 2007 to 2009



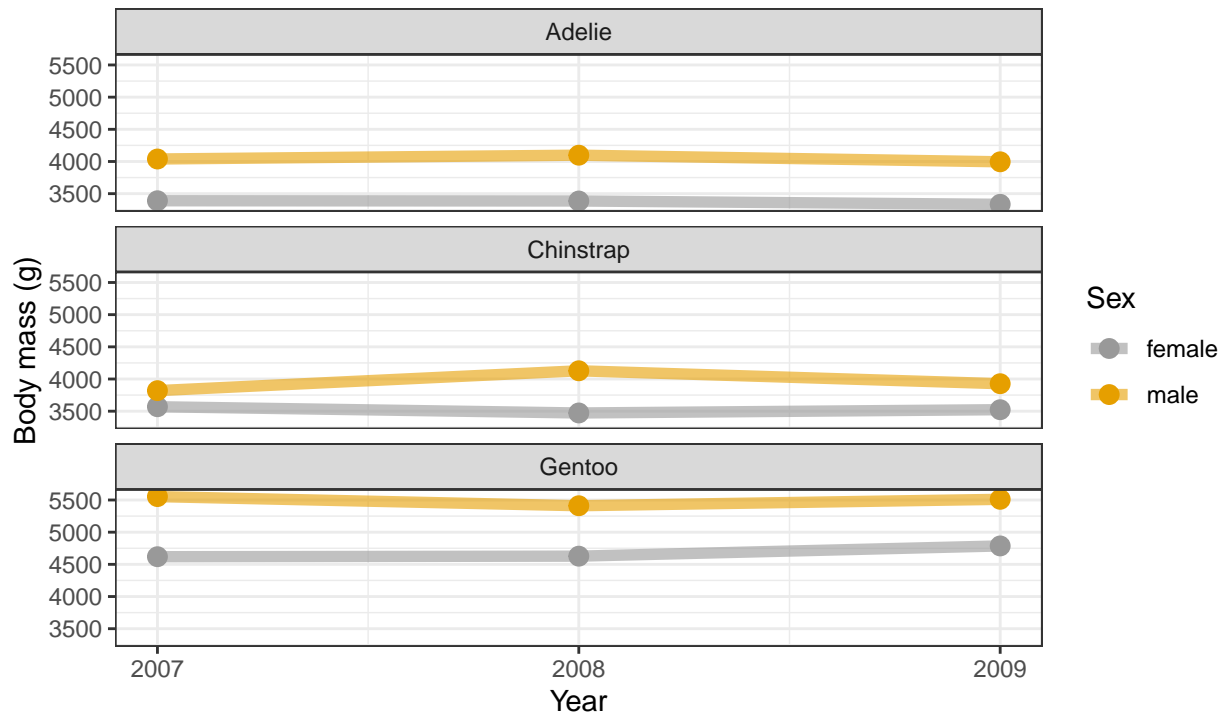
Source: <https://github.com/allisonhorst/palmerpenguins>

Slope charts

```
penguins %>%  
  remove_missing() %>%  
  group_by(year, species, sex) %>%  
  summarise(mean_bmg = mean(body_mass_g)) %>%  
  ggplot(aes(x = year, y = mean_bmg, group = sex,  
             color = sex)) +  
  facet_wrap(vars(species), nrow = 3) +  
  geom_line(alpha = 0.6, size = 2) +  
  geom_point(alpha = 1, size = 3) +  
  scale_x_continuous(breaks=c(2007, 2008, 2009)) +  
  labs(x = "Year",  
       y = "Body mass (g)",  
       color = "Sex",  
       title = "Slope chart",  
       subtitle = "Penguin's change in body mass from 2007 to 2009",  
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```


Slope chart

Penguin's change in body mass from 2007 to 2009



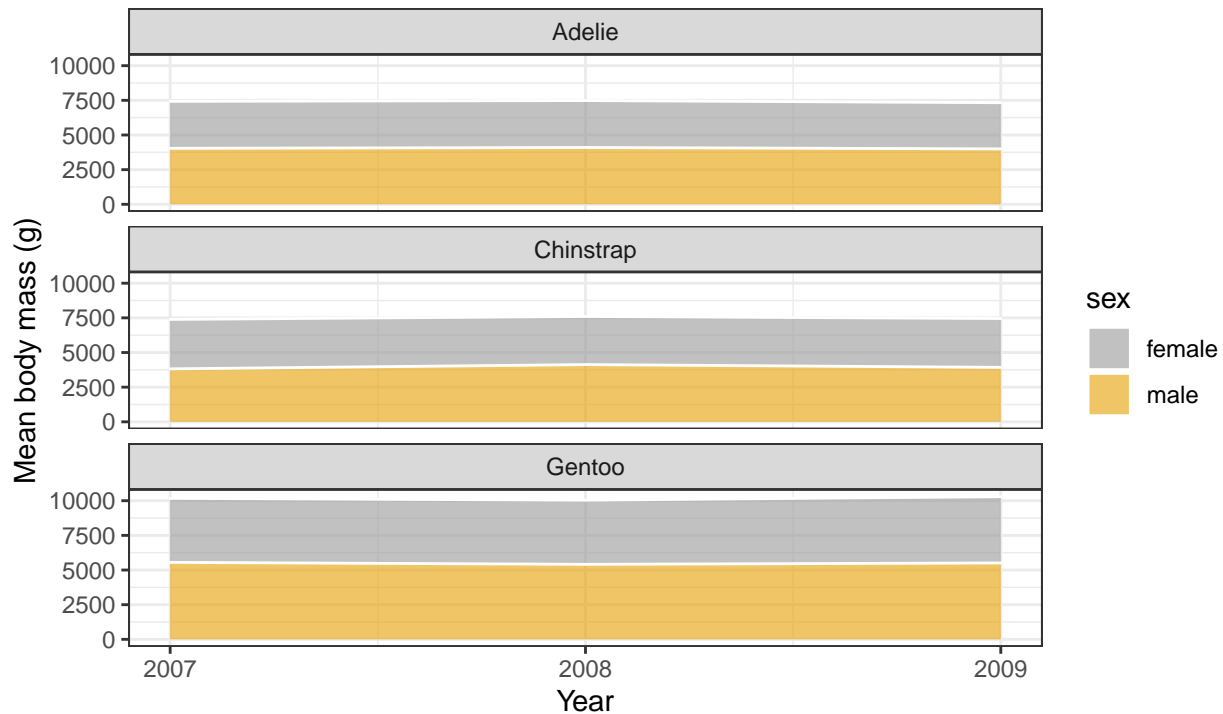
Source: <https://github.com/allisonhorst/palmerpenguins>

Stacked area charts

```
penguins %>%
  remove_missing() %>%
  group_by(year, species, sex) %>%
  summarise(mean_bmg = mean(body_mass_g)) %>%
  ggplot(aes(x = year, y = mean_bmg, fill = sex)) +
  facet_wrap(vars(species), nrow = 3) +
  geom_area(alpha = 0.6, size=.5, color = "white") +
  scale_x_continuous(breaks=c(2007, 2008, 2009)) +
  labs(x = "Year",
       y = "Mean body mass (g)",
       color = "Sex",
       title = "Stacked area chart",
       subtitle = "Penguin's change in body mass from 2007 to 2009",
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```

Stacked area chart

Penguin's change in body mass from 2007 to 2009



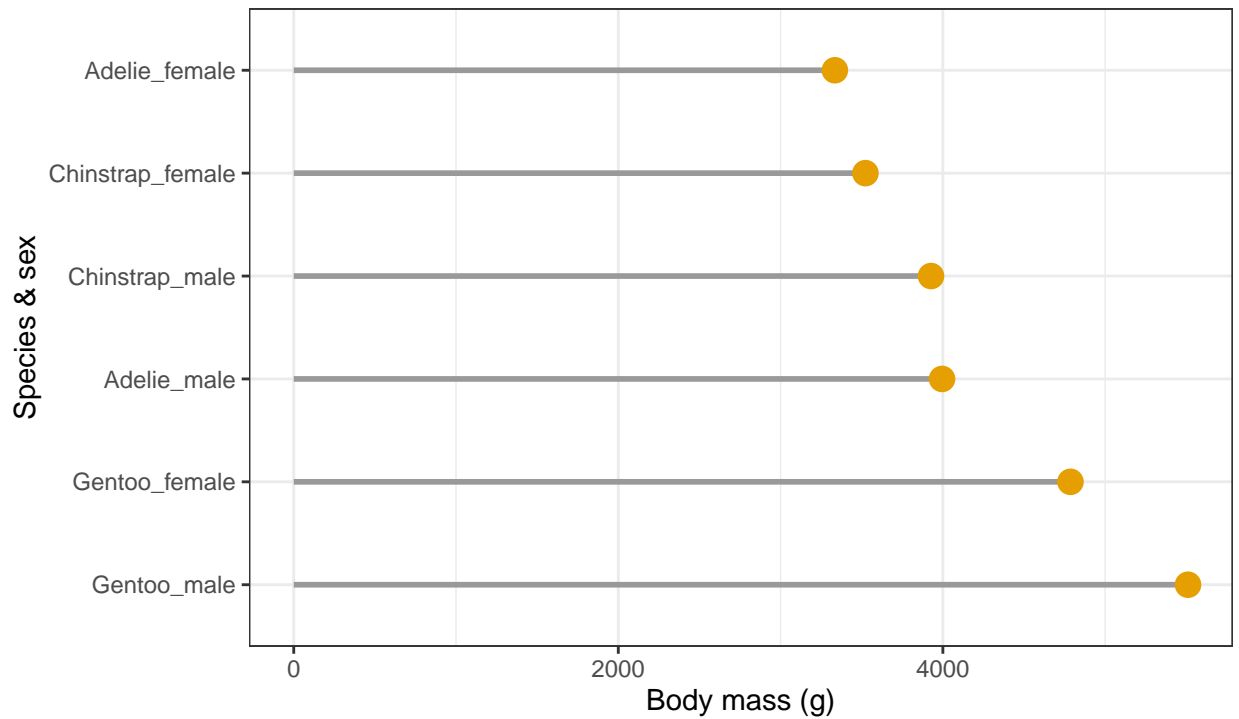
Source: <https://github.com/allisonhorst/palmerpenguins>

Lollipop chart

```
penguins %>%
  remove_missing() %>%
  group_by(year, species, sex) %>%
  summarise(mean_bmg = mean(body_mass_g)) %>%
  mutate(species_sex = paste(species, sex, sep = "_"),
         year = paste0("year_", year)) %>%
  spread(year, mean_bmg) %>%
  ggplot() +
  geom_segment(aes(x = reorder(species_sex, -year_2009), xend = reorder(species_sex, -year_2009),
                  y = 0, yend = year_2009),
             color = "#999999", size = 1) +
  geom_point(aes(x = reorder(species_sex, -year_2009), y = year_2009),
            size = 4, color = "#E69F00") +
  coord_flip() +
  labs(x = "Species & sex",
       y = "Body mass (g)",
       title = "Lollipop chart",
       subtitle = "Penguin's body mass in 2009",
       caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```

Lollipop chart

Penguin's body mass in 2009



Source: <https://github.com/allisonhorst/palmerpenguins>

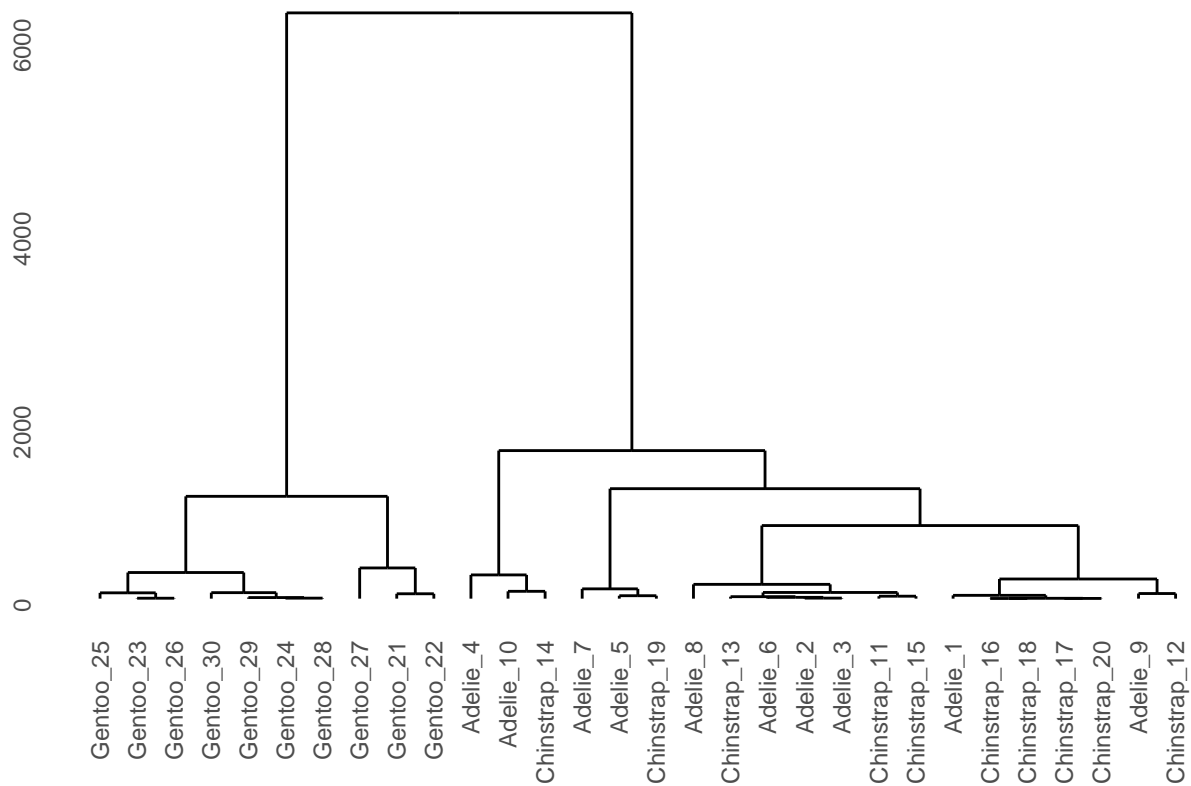
Dendrograms

```
library(ggdendro)
library(dendextend)
```

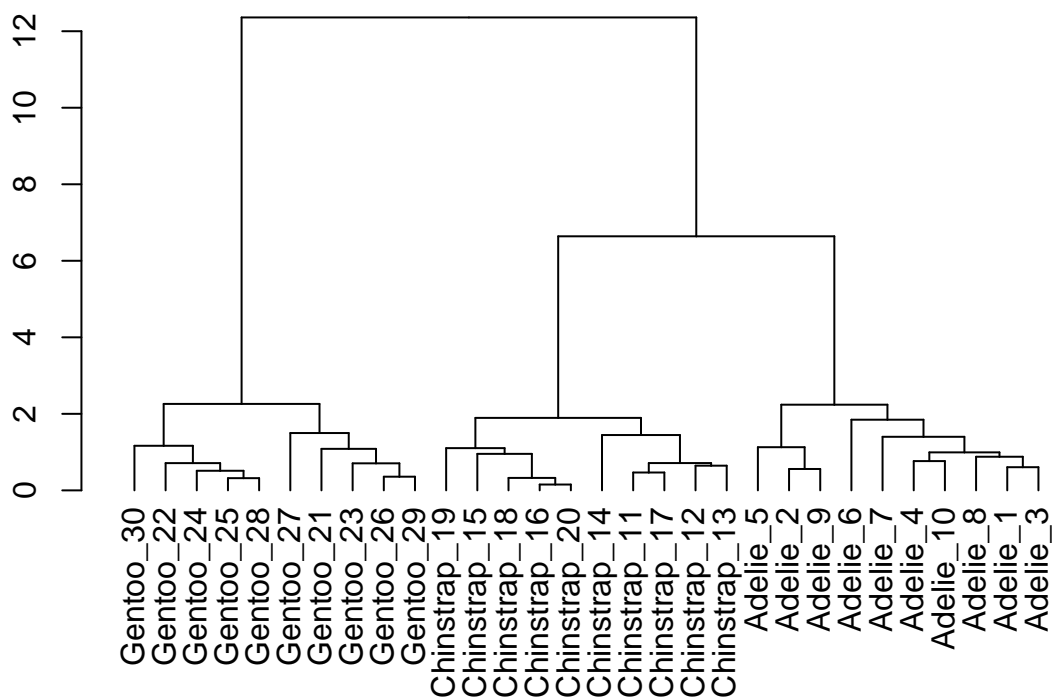
```
penguins_hist <- penguins %>%
  filter(sex == "male") %>%
  select(species, bill_length_mm, bill_depth_mm, flipper_length_mm, body_mass_g) %>%
  group_by(species) %>%
  sample_n(10) %>%
  as.data.frame()
rownames(penguins_hist) <- paste(penguins_hist$species, seq_len(nrow(penguins_hist)), sep = "_")

penguins_hist <- penguins_hist %>%
  select(-species) %>%
  remove_missing()

hc <- hclust(dist(penguins_hist, method = "euclidean"), method = "ward.D2")
ggdendrogram(hc)
```



```
# Create a dendrogram and plot it
penguins_hist %>%
  scale %>%
  dist(method = "euclidean") %>%
  hclust(method = "ward.D2") %>%
  as.dendrogram() %>%
  plot()
```



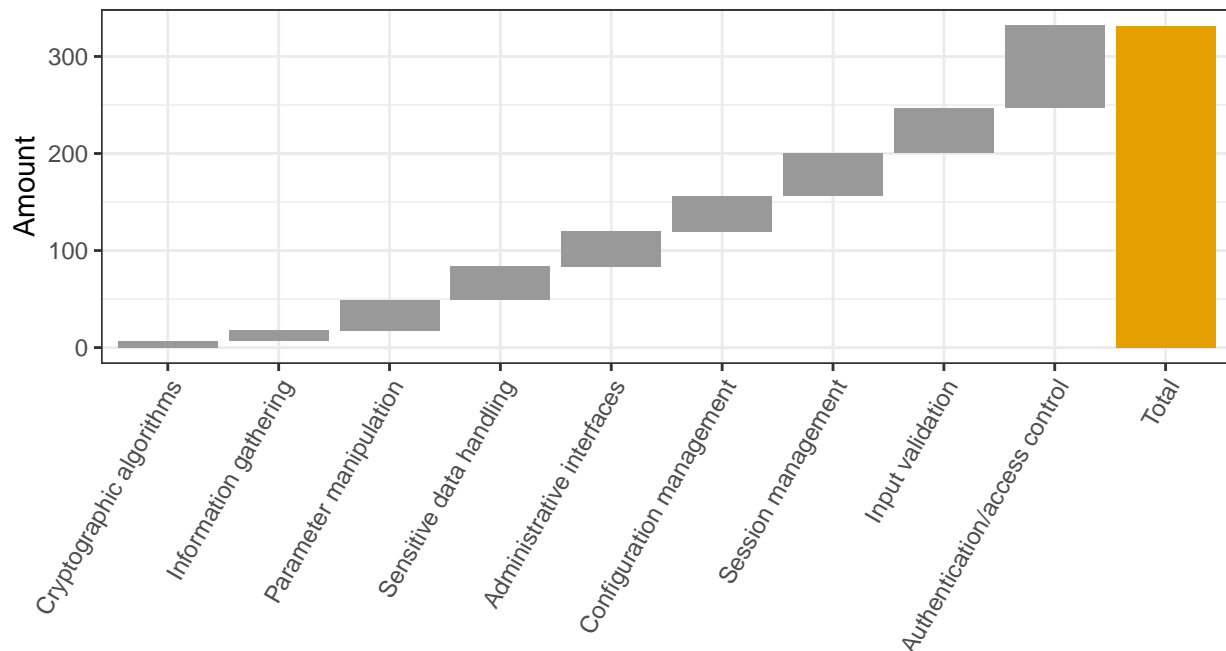
Waterfall charts

```
library(waterfall)
```

```
jaquith %>%  
  arrange(score) %>%  
  add_row(factor = "Total", score = sum(jaquith$score)) %>%  
  mutate(factor = factor(factor, levels = factor),  
         id = seq_along(score)) %>%  
  mutate(end = cumsum(score),  
         start = c(0, end[-length(end)]),  
         start = c(start[-length(start)], 0),  
         end = c(end[-length(end)], score[length(score)]),  
         gr_col = ifelse(factor == "Total", "Total", "Part")) %>%  
  ggplot(aes(x = factor, fill = gr_col)) +  
    geom_rect(aes(x = factor,  
                 xmin = id - 0.45, xmax = id + 0.45,  
                 ymin = end, ymax = start)) +  
    theme(axis.text.x = element_text(angle = 60, vjust = 1, hjust = 1),  
          legend.position = "none") +  
    labs(x = "",  
         y = "Amount",  
         title = "Waterfall chart",  
         subtitle = "Sample business-adjusted risk from Security Metrics",  
         caption = "Andrew Jaquith, Security Metrics: Replacing Fear, Uncertainty, and Doubt\n(Boston: Addison-Wesley Professional, 2007), 170-171.")
```

Waterfall chart

Sample business-adjusted risk from Security Metrics



Andrew Jaquith, Security Metrics: Replacing Fear, Uncertainty, and Doubt
(Boston: Addison-Wesley Professional, 2007), 170-171.

Biplots

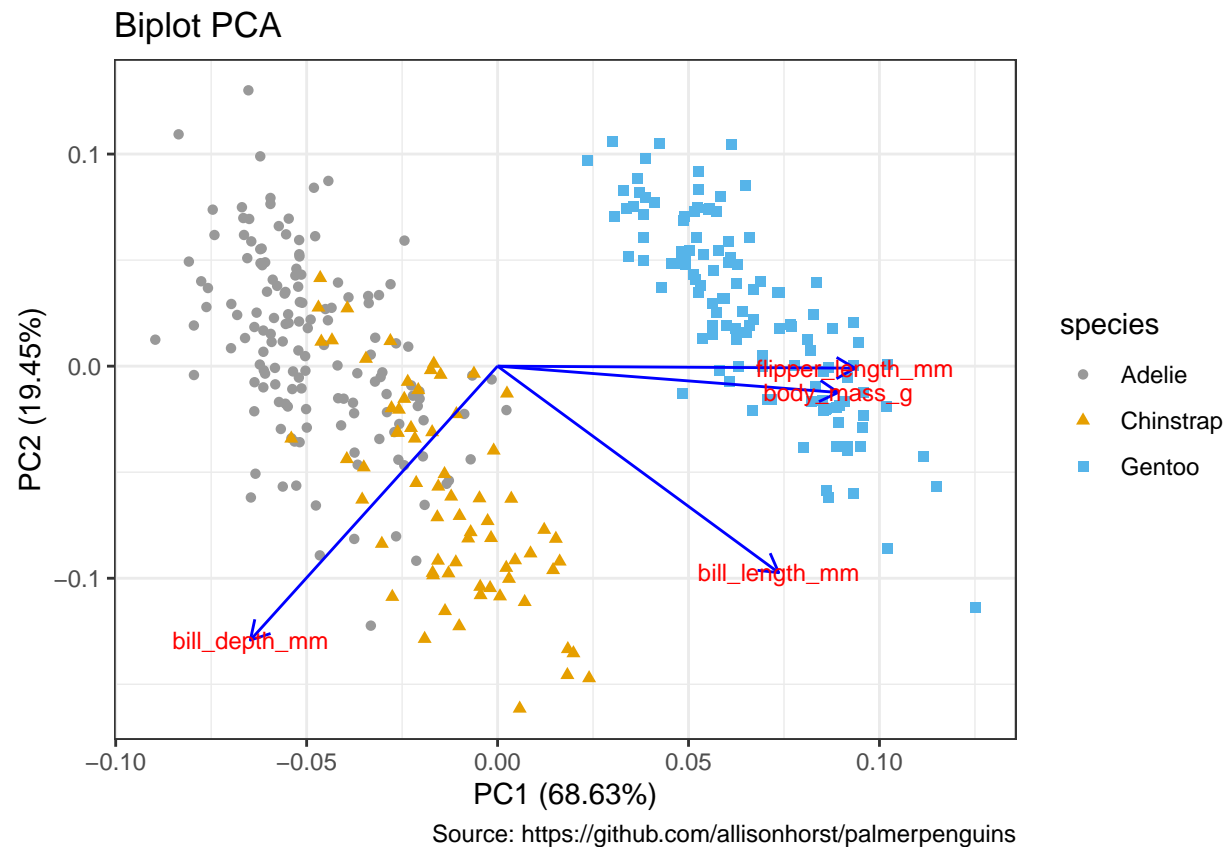
```
library(ggfortify)

penguins_prep <- penguins %>%
  remove_missing() %>%
  select(bill_length_mm:body_mass_g)

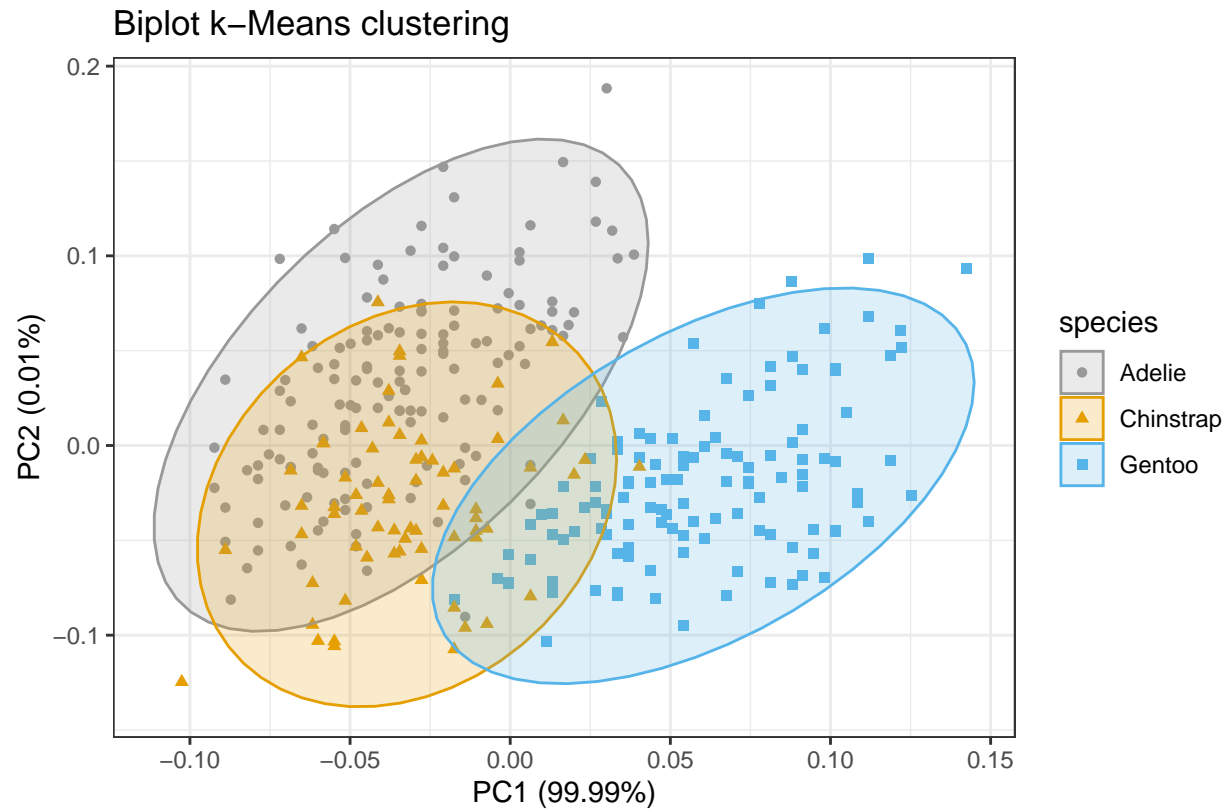
penguins_pca <- penguins_prep %>%
  prcomp(scale. = TRUE)

penguins_km <- penguins_prep %>%
  kmeans(3)

autoplot(penguins_pca,
  data = penguins %>% remove_missing(),
  colour = 'species',
  shape = 'species',
  loadings = TRUE,
  loadings.colour = 'blue',
  loadings.label = TRUE,
  loadings.label.size = 3) +
  scale_color_manual(values = cbp1) +
  scale_fill_manual(values = cbp1) +
  theme_bw() +
  labs(
    title = "Biplot PCA",
    caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```



```
autoplot(penguins_km,
  data = penguins %>% remove_missing(),
  colour = 'species',
  shape = 'species',
  frame = TRUE, frame.type = 'norm') +
  scale_color_manual(values = cbp1) +
  scale_fill_manual(values = cbp1) +
  theme_bw() +
  labs(
    title = "Biplot k-Means clustering",
    caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```



Source: <https://github.com/allisonhorst/palmerpenguins>

Radar charts, aka star chart, aka spider plot

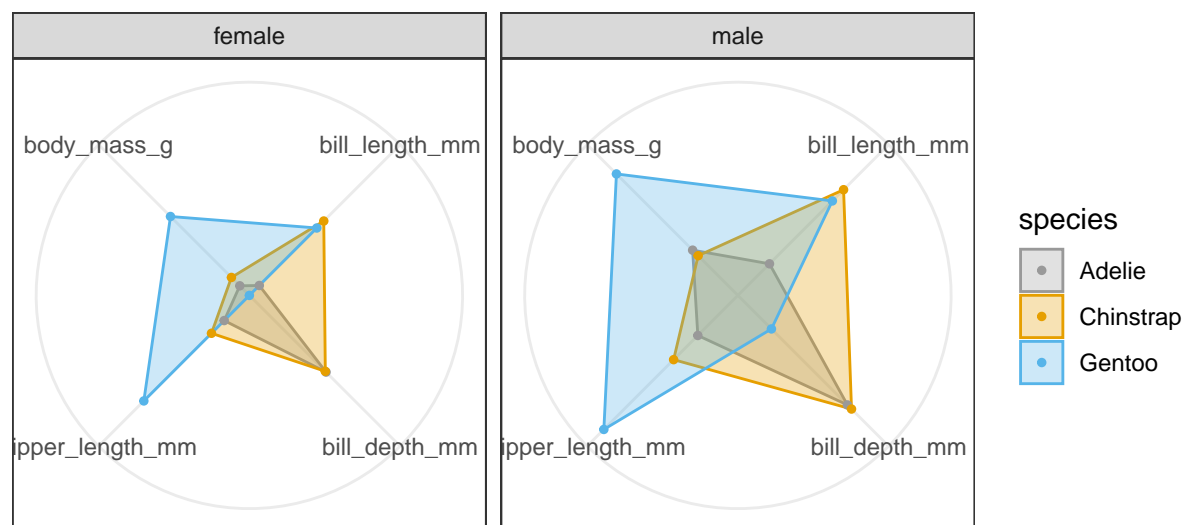
<https://www.data-to-viz.com/caveat/spider.html>

```
library(ggiraphExtra)
```

```
penguins %>%
  remove_missing() %>%
  select(-island, -year) %>%
  ggRadar(aes(x = c(bill_length_mm, bill_depth_mm, flipper_length_mm, body_mass_g),
    group = species,
    colour = sex, facet = sex),
    rescale = TRUE,
    size = 1, interactive = FALSE,
    use.label = TRUE) +
  scale_color_manual(values = cbp1) +
  scale_fill_manual(values = cbp1) +
  theme_bw() +
  scale_y_discrete(breaks = NULL) + # don't show ticks
  labs(
    title = "Radar/spider/star chart",
    subtitle = "Body mass of male & female penguins per species",
    caption = "Source: https://github.com/allisonhorst/palmerpenguins")
```


Radar/spider/star chart

Body mass of male & female penguins per species



Source: <https://github.com/allisonhorst/palmerpenguins>

```
devtools::session_info()
```

```
## - Session info -----
## setting value
## version R version 4.0.4 (2021-02-15)
## os      macOS Big Sur 10.16
## system  x86_64, darwin17.0
## ui      X11
## language (EN)
## collate en_US.UTF-8
## ctype   en_US.UTF-8
## tz      Europe/Berlin
## date    2021-04-21
##
## - Packages -----
## package      * version  date      lib source
## ash           1.0-15   2015-09-01 [2] CRAN (R 4.0.2)
## assertthat    0.2.1    2019-03-21 [2] CRAN (R 4.0.0)
## backports     1.2.1    2020-12-09 [2] CRAN (R 4.0.2)
## broom         0.7.5    2021-02-19 [2] CRAN (R 4.0.4)
## cachem        1.0.4    2021-02-13 [2] CRAN (R 4.0.2)
## callr         3.5.1    2020-10-13 [2] CRAN (R 4.0.2)
## cellranger    1.1.0    2016-07-27 [2] CRAN (R 4.0.0)
## cli           2.3.1    2021-02-23 [2] CRAN (R 4.0.4)
## colorspace    2.0-0    2020-11-11 [2] CRAN (R 4.0.2)
## crayon        1.4.1    2021-02-08 [2] CRAN (R 4.0.2)
## crosstalk     1.1.1    2021-01-12 [2] CRAN (R 4.0.2)
## data.table    1.14.0   2021-02-21 [2] CRAN (R 4.0.4)
## DBI           1.1.1    2021-01-15 [2] CRAN (R 4.0.2)
## dbplyr        2.1.0    2021-02-03 [2] CRAN (R 4.0.2)
```

## dendextend	* 1.14.0	2020-08-26	[2]	CRAN	(R 4.0.2)
## desc	1.3.0	2021-03-05	[2]	CRAN	(R 4.0.2)
## devtools	2.3.2	2020-09-18	[2]	CRAN	(R 4.0.2)
## digest	0.6.27	2020-10-24	[2]	CRAN	(R 4.0.2)
## dplyr	* 1.0.5	2021-03-05	[2]	CRAN	(R 4.0.2)
## ellipsis	0.3.1	2020-05-15	[2]	CRAN	(R 4.0.0)
## evaluate	0.14	2019-05-28	[2]	CRAN	(R 4.0.1)
## extrafont	0.17	2014-12-08	[2]	CRAN	(R 4.0.2)
## extrafontdb	1.0	2012-06-11	[2]	CRAN	(R 4.0.2)
## fansi	0.4.2	2021-01-15	[2]	CRAN	(R 4.0.2)
## farver	2.1.0	2021-02-28	[2]	CRAN	(R 4.0.2)
## fastmap	1.1.0	2021-01-25	[2]	CRAN	(R 4.0.2)
## forcats	* 0.5.1	2021-01-27	[2]	CRAN	(R 4.0.2)
## fs	1.5.0	2020-07-31	[2]	CRAN	(R 4.0.2)
## gdtools	0.2.3	2021-01-06	[2]	CRAN	(R 4.0.2)
## generics	0.1.0	2020-10-31	[2]	CRAN	(R 4.0.2)
## ggalluvial	* 0.12.3	2020-12-05	[2]	CRAN	(R 4.0.2)
## ggalt	* 0.4.0	2017-02-15	[2]	CRAN	(R 4.0.2)
## ggdendro	* 0.1.22	2020-09-13	[2]	CRAN	(R 4.0.2)
## ggExtra	* 0.9	2019-08-27	[2]	CRAN	(R 4.0.2)
## ggfittext	0.9.1	2021-01-30	[2]	CRAN	(R 4.0.2)
## ggfortify	* 0.4.11	2020-10-02	[2]	CRAN	(R 4.0.2)
## ggiraph	0.7.8	2020-07-01	[2]	CRAN	(R 4.0.2)
## ggiraphExtra	* 0.3.0	2020-10-06	[2]	CRAN	(R 4.0.2)
## ggplot2	* 3.3.3	2020-12-30	[2]	CRAN	(R 4.0.2)
## glue	1.4.2	2020-08-27	[2]	CRAN	(R 4.0.2)
## gridExtra	2.3	2017-09-09	[2]	CRAN	(R 4.0.2)
## gtable	0.3.0	2019-03-25	[2]	CRAN	(R 4.0.0)
## haven	2.3.1	2020-06-01	[2]	CRAN	(R 4.0.2)
## highr	0.8	2019-03-20	[2]	CRAN	(R 4.0.0)
## hms	1.0.0	2021-01-13	[2]	CRAN	(R 4.0.2)
## htmltools	0.5.1.1	2021-01-22	[2]	CRAN	(R 4.0.2)
## htmlwidgets	1.5.3	2020-12-10	[2]	CRAN	(R 4.0.2)
## httpuv	1.5.5	2021-01-13	[2]	CRAN	(R 4.0.2)
## httr	1.4.2	2020-07-20	[2]	CRAN	(R 4.0.2)
## insight	0.13.1	2021-02-22	[2]	CRAN	(R 4.0.4)
## jsonlite	1.7.2	2020-12-09	[2]	CRAN	(R 4.0.2)
## KernSmooth	2.23-18	2020-10-29	[2]	CRAN	(R 4.0.4)
## knitr	1.31	2021-01-27	[2]	CRAN	(R 4.0.2)
## labeling	0.4.2	2020-10-20	[2]	CRAN	(R 4.0.2)
## later	1.1.0.1	2020-06-05	[2]	CRAN	(R 4.0.2)
## lattice	* 0.20-41	2020-04-02	[2]	CRAN	(R 4.0.4)
## lazyeval	0.2.2	2019-03-15	[2]	CRAN	(R 4.0.0)
## lifecycle	1.0.0	2021-02-15	[2]	CRAN	(R 4.0.2)
## lubridate	1.7.10	2021-02-26	[2]	CRAN	(R 4.0.2)
## magrittr	2.0.1	2020-11-17	[2]	CRAN	(R 4.0.2)
## maps	3.3.0	2018-04-03	[2]	CRAN	(R 4.0.2)
## MASS	7.3-53.1	2021-02-12	[2]	CRAN	(R 4.0.2)
## Matrix	1.3-2	2021-01-06	[2]	CRAN	(R 4.0.4)
## memoise	2.0.0	2021-01-26	[2]	CRAN	(R 4.0.2)
## mgcv	1.8-34	2021-02-16	[2]	CRAN	(R 4.0.2)
## mime	0.10	2021-02-13	[2]	CRAN	(R 4.0.2)
## miniUI	0.1.1.1	2018-05-18	[2]	CRAN	(R 4.0.0)
## modelr	0.1.8	2020-05-19	[2]	CRAN	(R 4.0.2)

##	munsell	0.5.0	2018-06-12	[2]	CRAN	(R 4.0.0)
##	mycor	0.1.1	2018-04-10	[2]	CRAN	(R 4.0.2)
##	nlme	3.1-152	2021-02-04	[2]	CRAN	(R 4.0.4)
##	palmerpenguins	* 0.1.0	2020-07-23	[2]	CRAN	(R 4.0.2)
##	pillar	1.5.1	2021-03-05	[2]	CRAN	(R 4.0.2)
##	pkgbuild	1.2.0	2020-12-15	[2]	CRAN	(R 4.0.2)
##	pkgconfig	2.0.3	2019-09-22	[2]	CRAN	(R 4.0.0)
##	pkgload	1.2.0	2021-02-23	[2]	CRAN	(R 4.0.4)
##	plotly	* 4.9.3	2021-01-10	[2]	CRAN	(R 4.0.2)
##	plotrix	* 3.8-1	2021-01-21	[2]	CRAN	(R 4.0.2)
##	plyr	1.8.6	2020-03-03	[2]	CRAN	(R 4.0.0)
##	ppcor	1.1	2015-12-03	[2]	CRAN	(R 4.0.2)
##	prettyunits	1.1.1	2020-01-24	[2]	CRAN	(R 4.0.0)
##	processx	3.4.5	2020-11-30	[2]	CRAN	(R 4.0.2)
##	proj4	1.0-10.1	2021-01-26	[2]	CRAN	(R 4.0.2)
##	promises	1.2.0.1	2021-02-11	[2]	CRAN	(R 4.0.2)
##	ps	1.6.0	2021-02-28	[2]	CRAN	(R 4.0.2)
##	purrr	* 0.3.4	2020-04-17	[2]	CRAN	(R 4.0.0)
##	R6	2.5.0	2020-10-28	[2]	CRAN	(R 4.0.2)
##	ragg	* 1.1.1	2021-02-25	[2]	CRAN	(R 4.0.2)
##	RColorBrewer	1.1-2	2014-12-07	[2]	CRAN	(R 4.0.0)
##	Rcpp	1.0.6	2021-01-15	[2]	CRAN	(R 4.0.2)
##	readr	* 1.4.0	2020-10-05	[2]	CRAN	(R 4.0.2)
##	readxl	1.3.1	2019-03-13	[1]	CRAN	(R 4.0.2)
##	remotes	2.2.0	2020-07-21	[2]	CRAN	(R 4.0.2)
##	reprex	1.0.0	2021-01-27	[2]	CRAN	(R 4.0.2)
##	reshape2	1.4.4	2020-04-09	[2]	CRAN	(R 4.0.0)
##	rlang	0.4.10	2020-12-30	[2]	CRAN	(R 4.0.2)
##	rmarkdown	2.7	2021-02-19	[2]	CRAN	(R 4.0.4)
##	rprojroot	2.0.2	2020-11-15	[2]	CRAN	(R 4.0.2)
##	rstudioapi	0.13	2020-11-12	[2]	CRAN	(R 4.0.2)
##	Rttf2pt1	1.3.8	2020-01-10	[2]	CRAN	(R 4.0.2)
##	rvest	1.0.0	2021-03-09	[2]	CRAN	(R 4.0.2)
##	scales	1.1.1	2020-05-11	[2]	CRAN	(R 4.0.0)
##	sessioninfo	1.1.1	2018-11-05	[2]	CRAN	(R 4.0.0)
##	shiny	1.6.0	2021-01-25	[2]	CRAN	(R 4.0.2)
##	sjlabelled	1.1.7	2020-09-24	[2]	CRAN	(R 4.0.2)
##	sjmisc	2.8.6	2021-01-07	[2]	CRAN	(R 4.0.2)
##	stringi	1.5.3	2020-09-09	[2]	CRAN	(R 4.0.2)
##	stringr	* 1.4.0	2019-02-10	[2]	CRAN	(R 4.0.0)
##	systemfonts	1.0.1	2021-02-09	[2]	CRAN	(R 4.0.2)
##	testthat	3.0.2	2021-02-14	[2]	CRAN	(R 4.0.2)
##	textshaping	0.3.2	2021-03-10	[2]	CRAN	(R 4.0.2)
##	tibble	* 3.1.0	2021-02-25	[2]	CRAN	(R 4.0.2)
##	tidyr	* 1.1.3	2021-03-03	[2]	CRAN	(R 4.0.2)
##	tidyselect	1.1.0	2020-05-11	[2]	CRAN	(R 4.0.0)
##	tidyverse	* 1.3.0	2019-11-21	[2]	CRAN	(R 4.0.0)
##	treemapify	* 2.5.5	2021-01-08	[2]	CRAN	(R 4.0.2)
##	usethis	2.0.1	2021-02-10	[2]	CRAN	(R 4.0.2)
##	utf8	1.2.1	2021-03-12	[2]	CRAN	(R 4.0.4)
##	uuid	0.1-4	2020-02-26	[2]	CRAN	(R 4.0.2)
##	vctrs	0.3.6	2020-12-17	[2]	CRAN	(R 4.0.2)
##	viridis	0.5.1	2018-03-29	[2]	CRAN	(R 4.0.2)
##	viridisLite	0.3.0	2018-02-01	[2]	CRAN	(R 4.0.0)

```
## waterfall      * 1.0.2    2016-04-03 [2] CRAN (R 4.0.2)
## webshot        0.5.2    2019-11-22 [2] CRAN (R 4.0.2)
## withr          2.4.1    2021-01-26 [2] CRAN (R 4.0.2)
## xfun           0.22     2021-03-11 [2] CRAN (R 4.0.2)
## xml2           1.3.2    2020-04-23 [2] CRAN (R 4.0.0)
## xtable         1.8-4    2019-04-21 [2] CRAN (R 4.0.0)
## yaml           2.2.1    2020-02-01 [2] CRAN (R 4.0.0)
##
## [1] /Users/shiringlander/Library/R/4.0/library
## [2] /Library/Frameworks/R.framework/Versions/4.0/Resources/library
```