Data Visualization

Chinazom

2024-04-04

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.3
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(modeldata)
```

```
?crickets
```

```
## starting httpd help server ... done
```

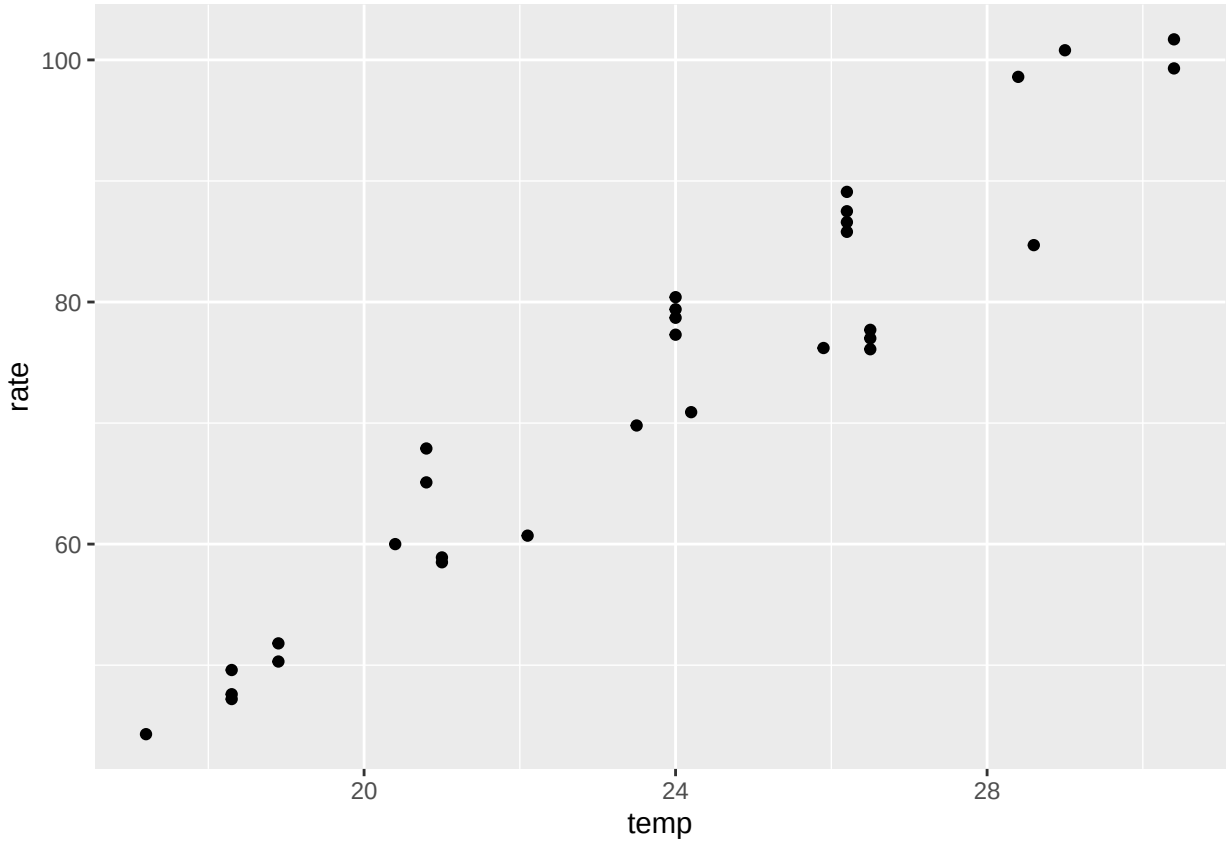
```
View(crickets)
```

The basics

Creating a chart for temp and rate

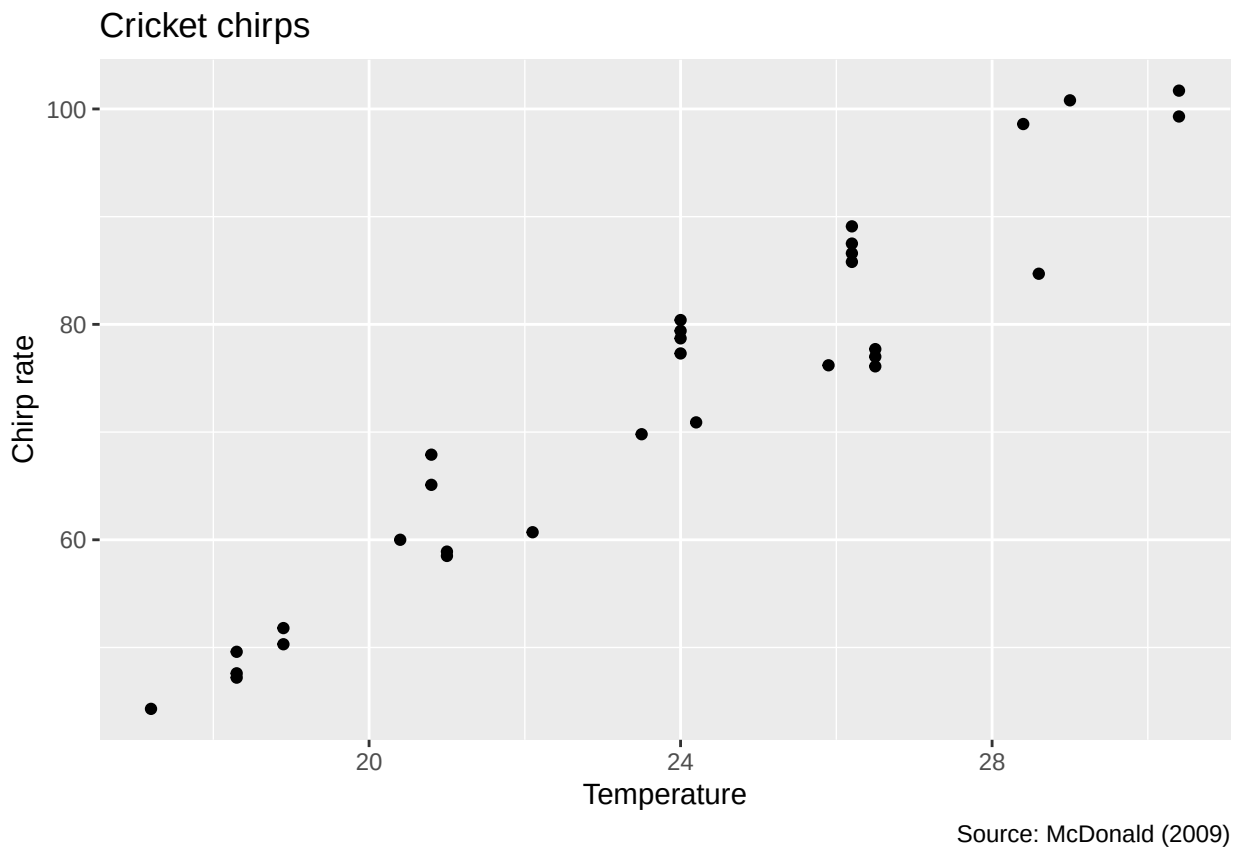
geom_point known as scatter plot is good for two quantitative variable

```
ggplot(crickets, aes(x = temp,
                     y = rate)) +
  geom_point()
```



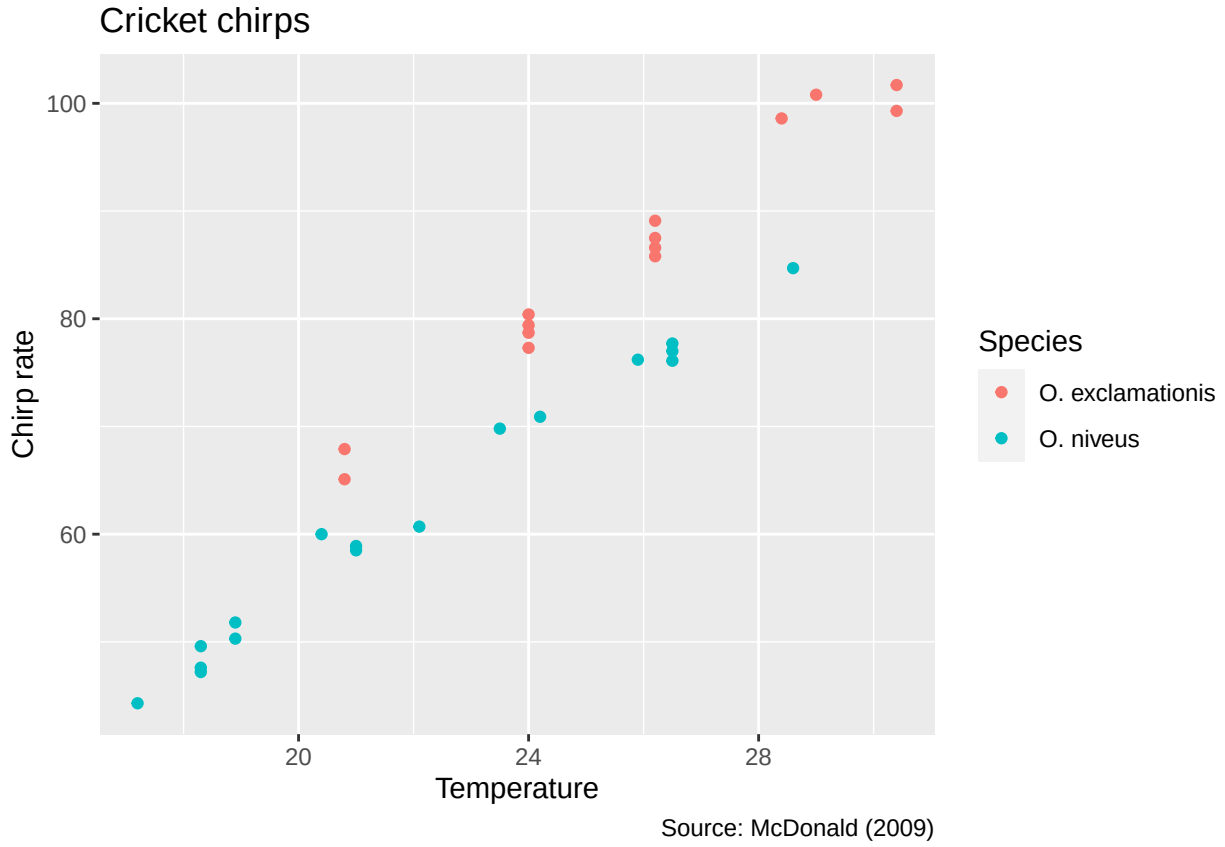
To label the chart

```
ggplot(crickets, aes(x = temp,
                     y = rate)) +
  geom_point() +
  labs(x = "Temperature",
       y = "Chirp rate",
       title = "Cricket chirps",
       caption = "Source: McDonald (2009)")
```



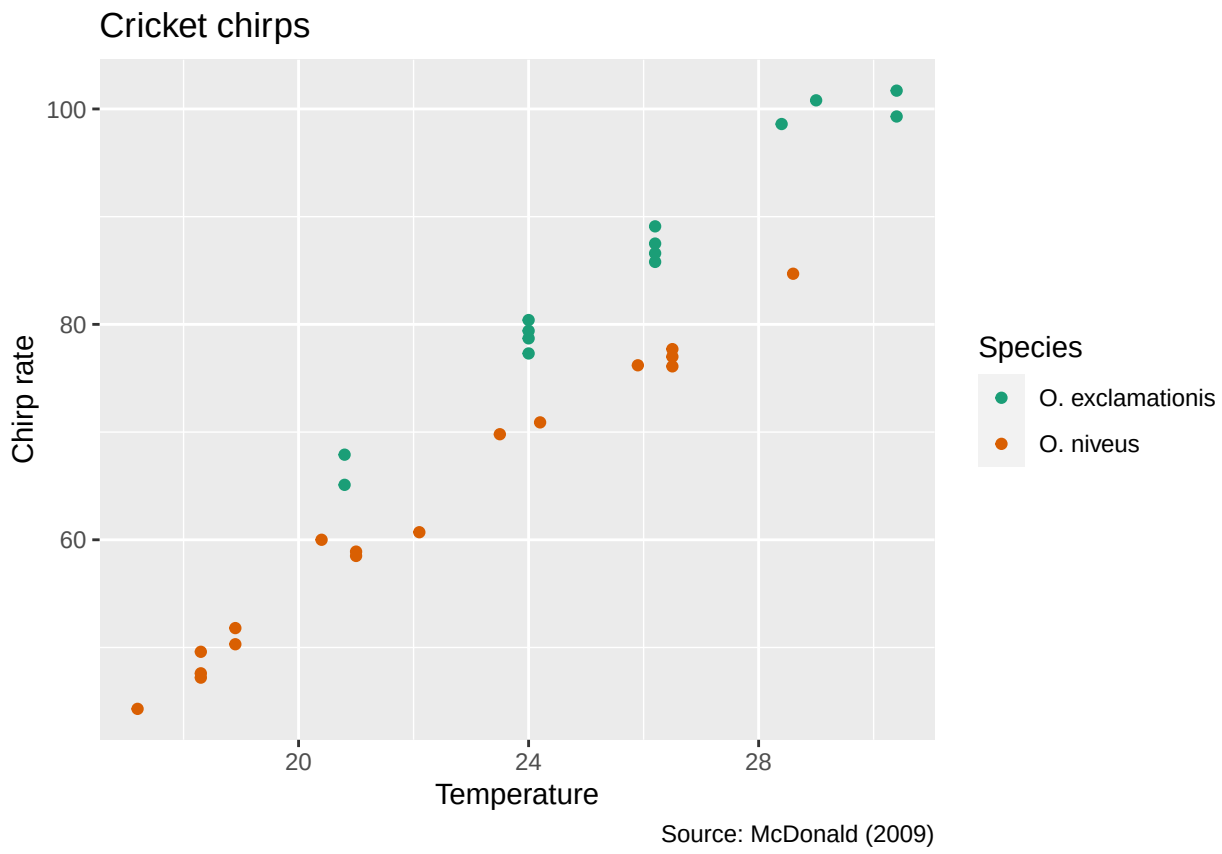

```
ggplot(crickets, aes(x = temp,
                     y = rate,
                     color = species)) +
  geom_point() +
  labs(x = "Temperature",
       y = "Chirp rate",
       color = "Species",
       title = "Cricket chirps",
       caption = "Source: McDonald (2009)")
```

To change the color of the charts based on the different species, you add color = species



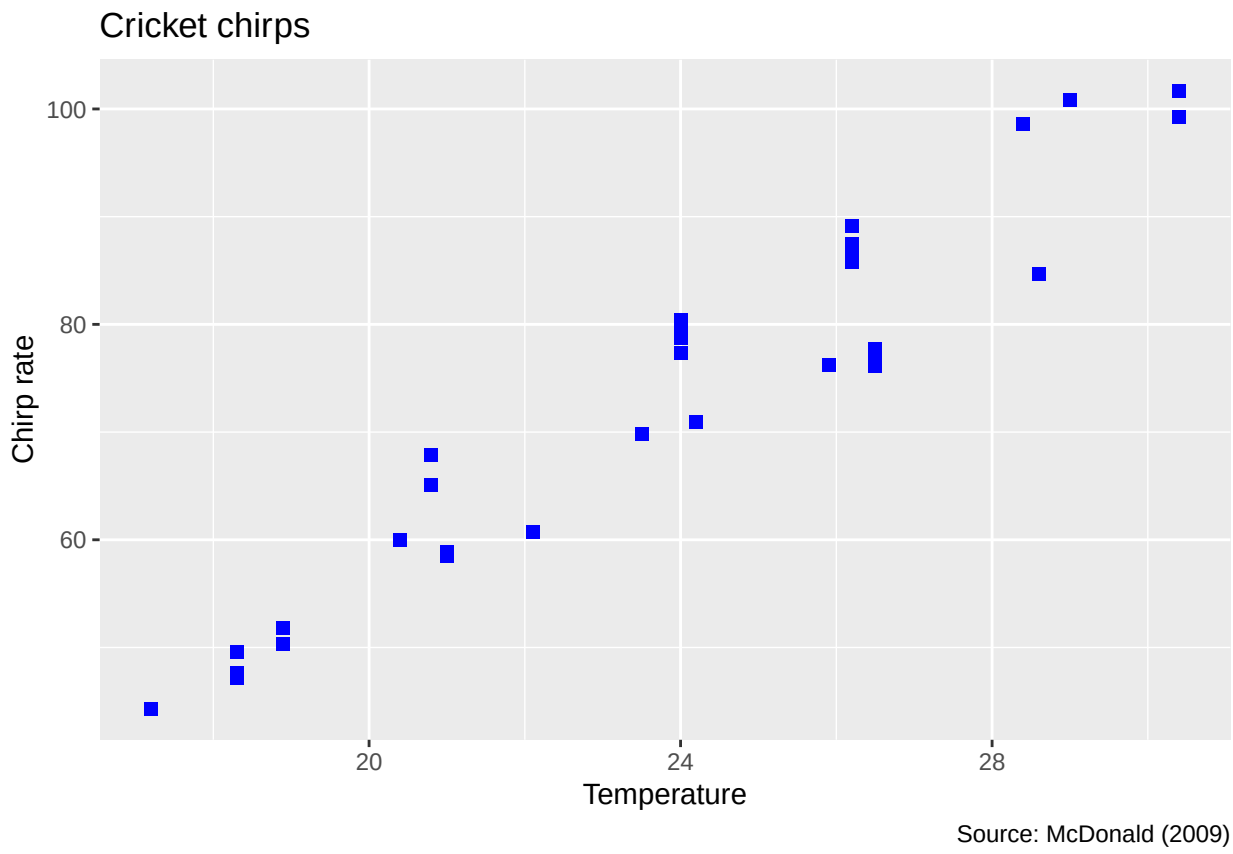
To get a darker shade of the color using scale_color_brewer

```
ggplot(crickets, aes(x = temp,
                     y = rate,
                     color = species)) +
  geom_point() +
  labs(x = "Temperature",
       y = "Chirp rate",
       color = "Species",
       title = "Cricket chirps",
       caption = "Source: McDonald (2009)") +
       scale_color_brewer(palette = "Dark2")
```



Modifiying basic properties of the plot

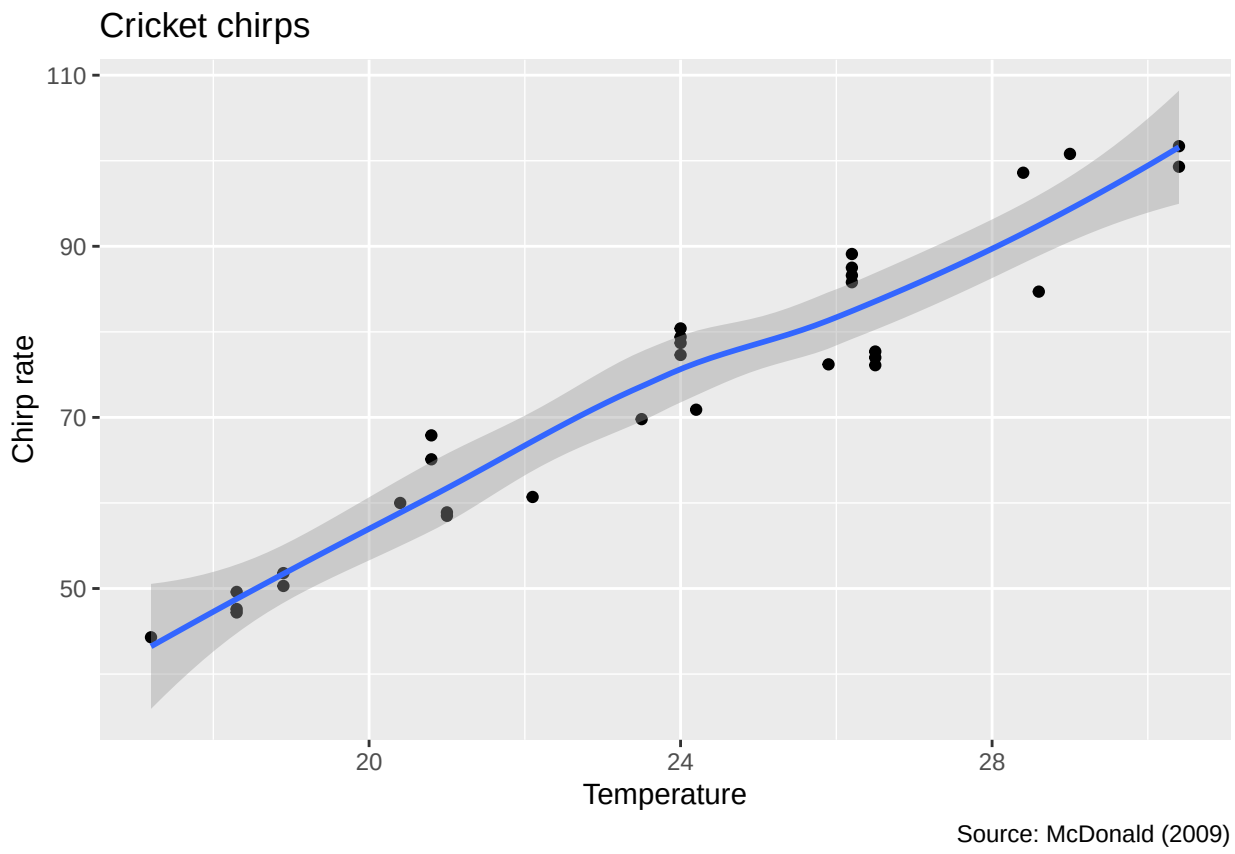
```
ggplot(crickets, aes(x = temp,
                     y = rate)) +
  geom_point(color = "blue",
             size = 2,
             shape = "square") +
  labs(x = "Temperature",
       y = "Chirp rate",
       title = "Cricket chirps",
       caption = "Source: McDonald (2009)")
```



Adding another layer geom_smooth to show the direction of the relationship

```
ggplot(crickets, aes(x = temp,
                     y = rate)) +
  geom_point() +
  geom_smooth() +
  labs(x = "Temperature",
       y = "Chirp rate",
       title = "Cricket chirps",
       caption = "Source: McDonald (2009)")
```

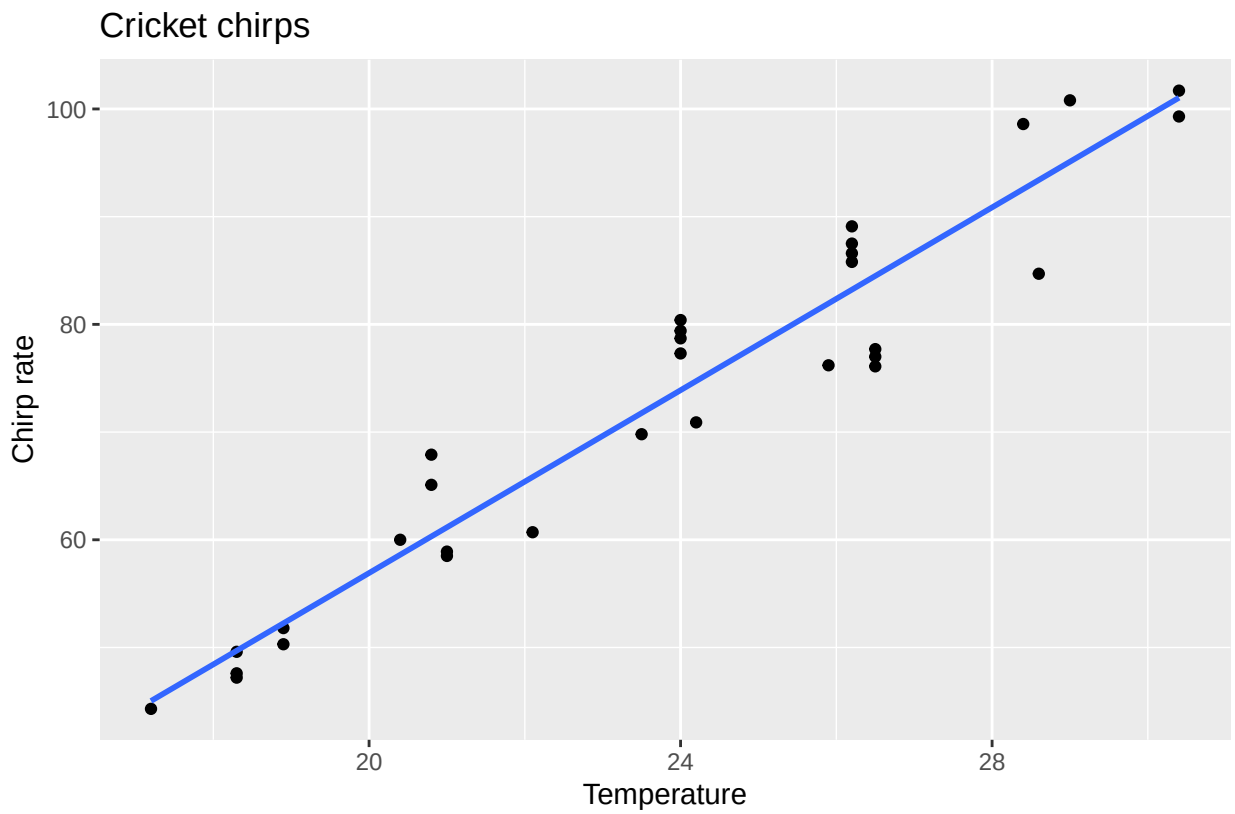
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



To modify the geom_smooth

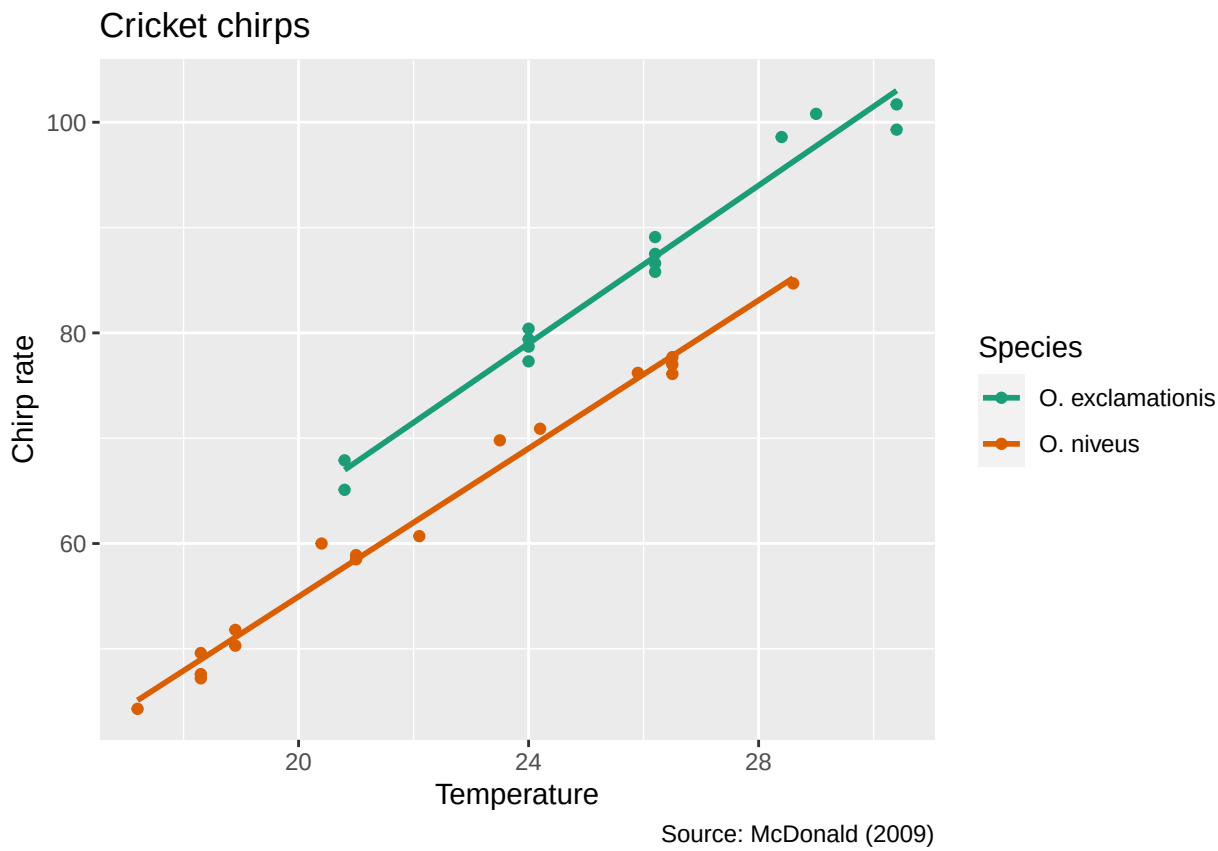
```
ggplot(crickets, aes(x = temp,
                     y = rate)) +
  geom_point() +
  geom_smooth(method = "lm", # to make the line linear
              se = FALSE) + # se = false takes away the shadow from the line
  labs(x = "Temperature",
       y = "Chirp rate",
       title = "Cricket chirps",
       caption = "Source: McDonald (2009)")
```

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



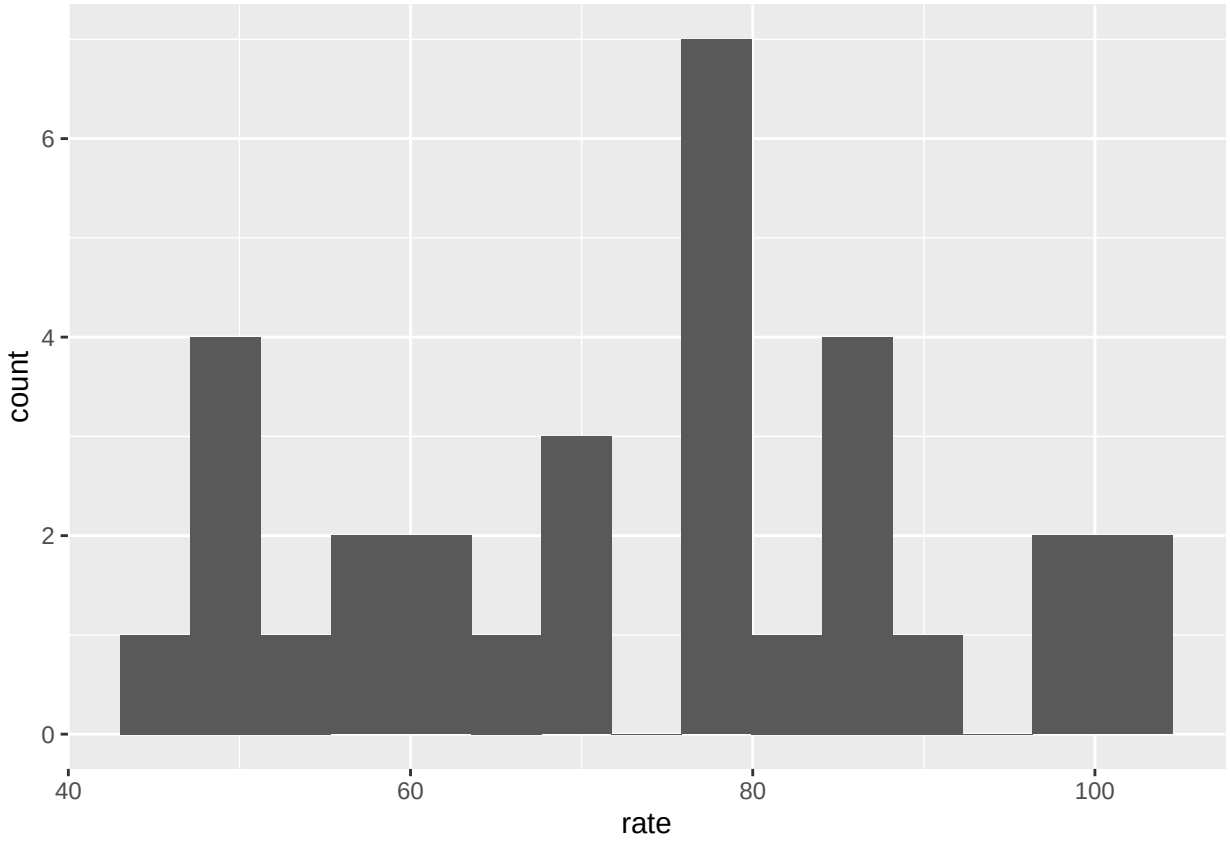

```
ggplot(crickets, aes(x = temp,
                     y = rate,
                     color = species)) + # to differentiate between the different speciess
  geom_point() +
  geom_smooth(method = "lm",
              se = FALSE) +
  labs(x = "Temperature",
       y = "Chirp rate",
       color = "Species",
       title = "Cricket chirps",
       caption = "Source: McDonald (2009)") +
  scale_color_brewer(palette = "Dark2") # to make the colors darker
```

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



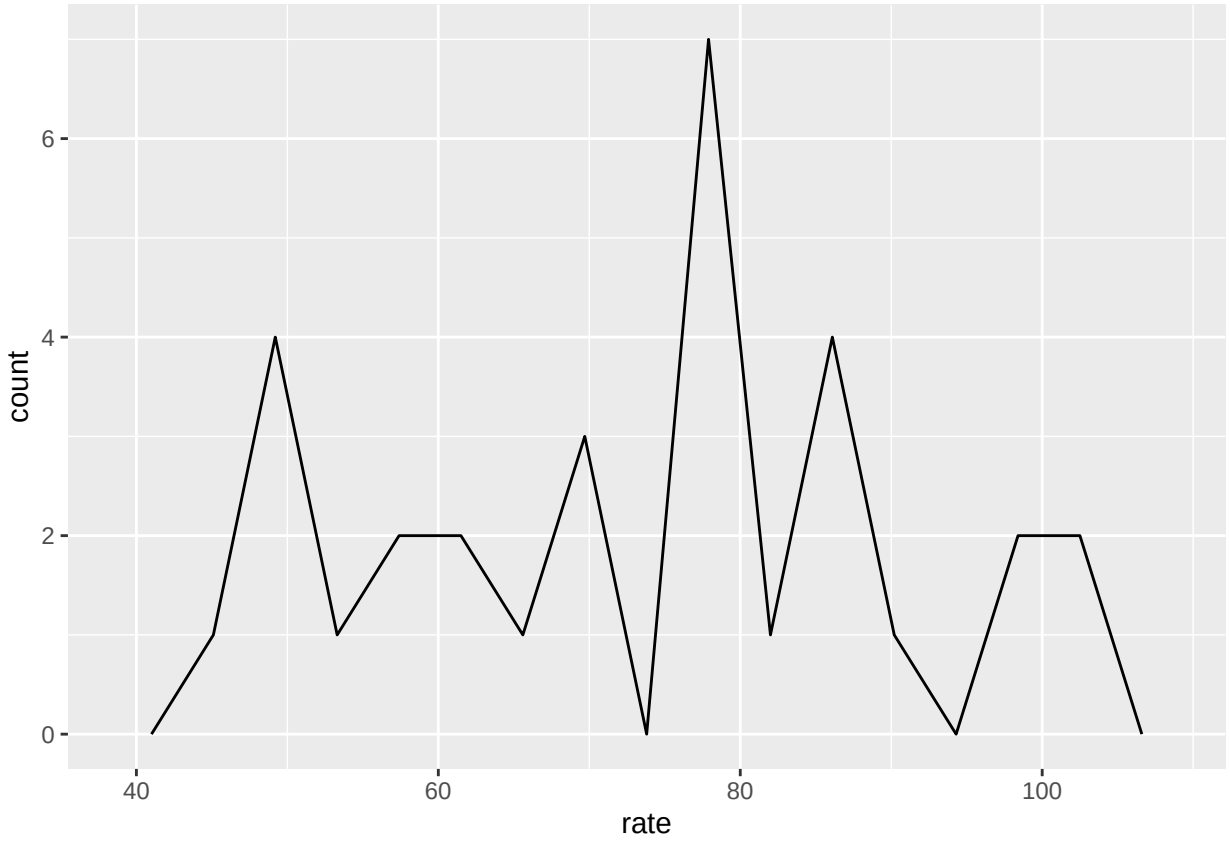
Other plots ### Histogram- mostly use for one quantitative variable

```
ggplot(crickets, aes(x = rate)) +
  geom_histogram(bins = 15) #  bin represents a group or a segment of the data, interval into which the
```



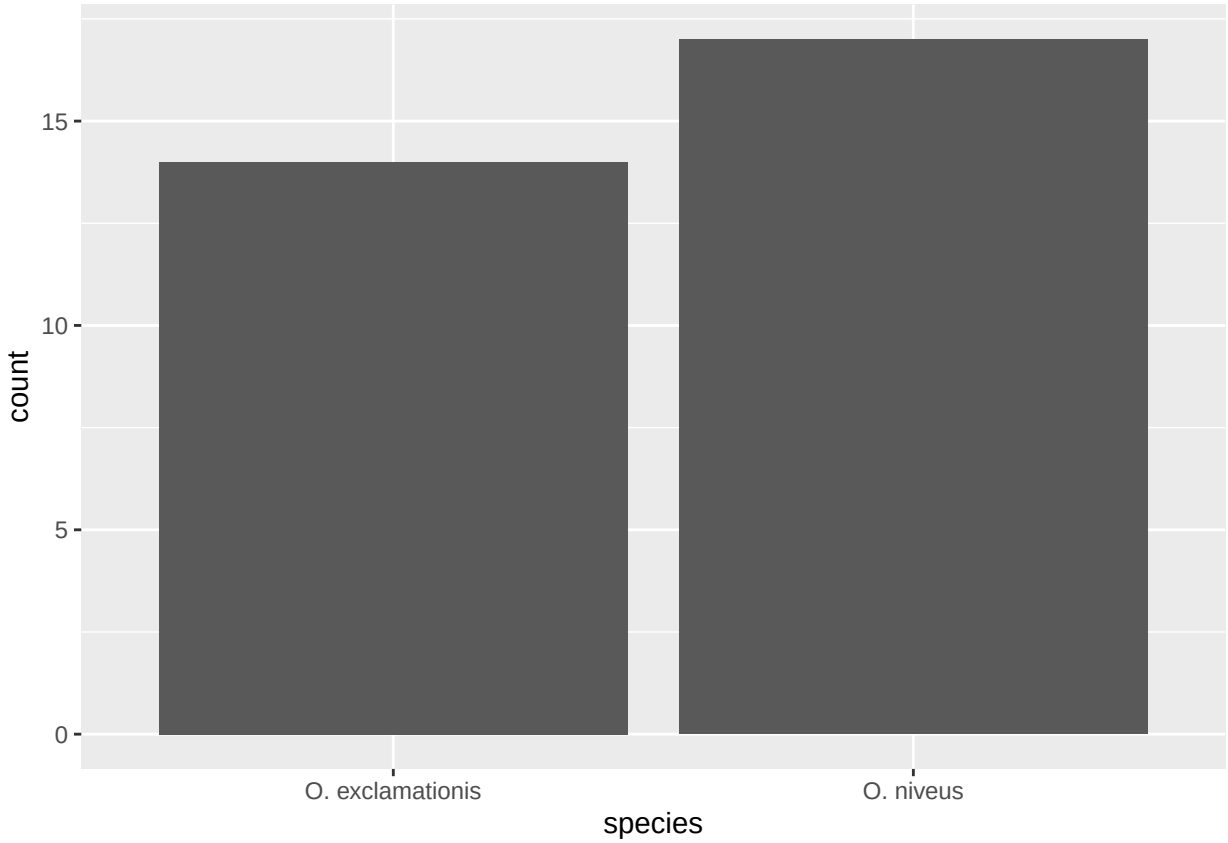
Geom_frepoly - uses lines to represent the frequency distribution of a continuous variable.

```
ggplot(crickets, aes(x = rate)) +
  geom_freqpoly(bins = 15)
```



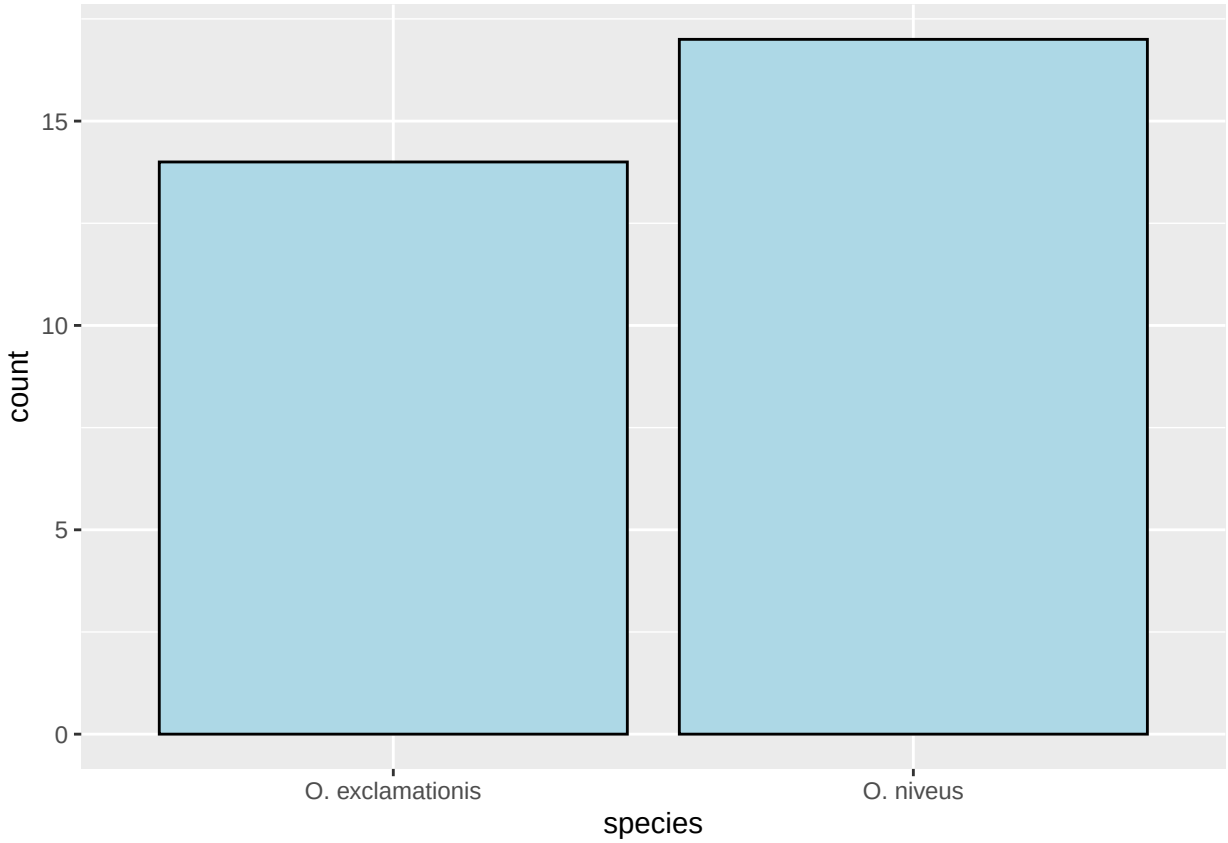
Bar chart mostly used for one categorical variable

```
ggplot(crickets, aes(x = species)) +
  geom_bar()
```



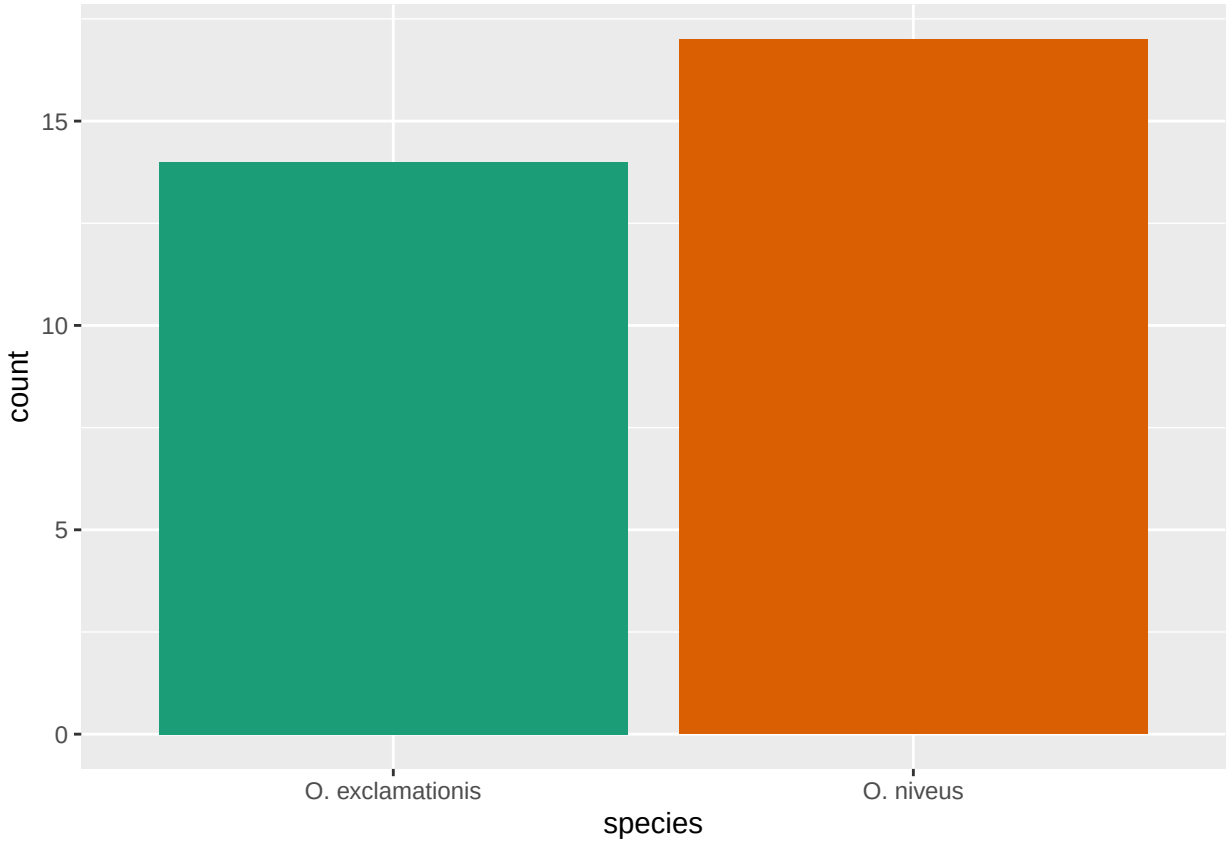
Modification of the Bar chart output #### To change the outer and inner color of the bar

```
ggplot(crickets, aes(x = species)) +
  geom_bar (color = "black", # color function works for the borders of the chart
fill = "lightblue") # fill for the inside
```



To give the values in a variable different colors - here species

```
ggplot(crickets, aes(x = species,
                     fill = species)) +
  geom_bar(show.legend = FALSE) + # to remove the legend from the chart
  scale_fill_brewer(palette = "Dark2") # to make the fill darker
```

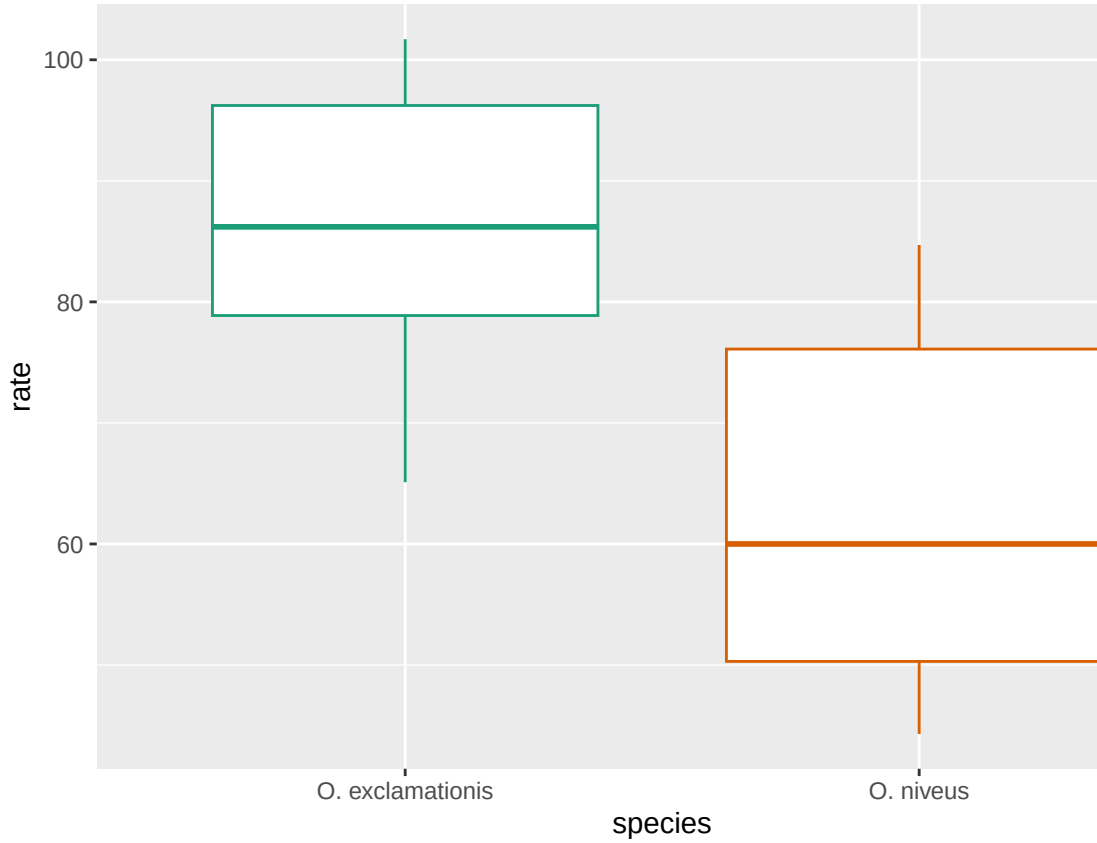


Geom_boxplot is good for one categorical and one quantitative variables

Boxplot is a visual representation of your 5 number summary; minimum - the line below

```
ggplot(crickets, aes(x = species,
                     y = rate,
                     color = species)) +
  geom_boxplot(show.legend = FALSE) +
  scale_color_brewer(palette = "Dark2") # To make the color darker
```

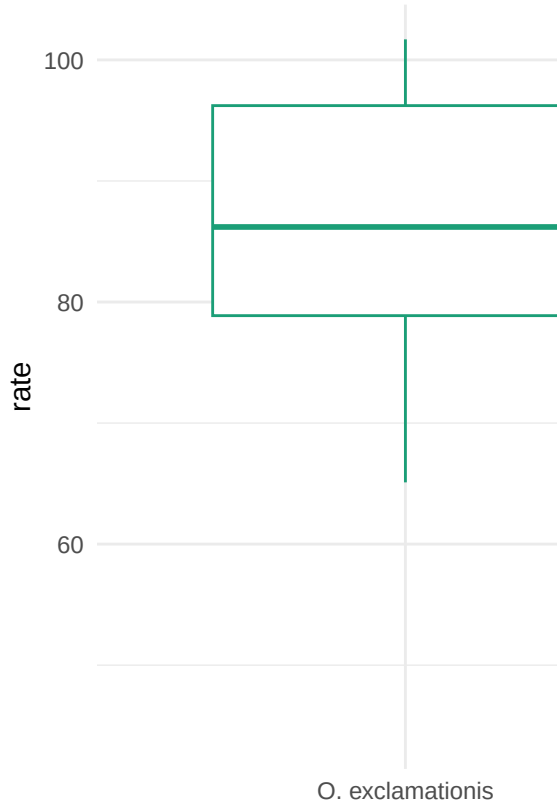
1st quartile, the 2nd quartile which is the medium (the line in-between), the 3rd quartile and



the max - the line above

Theme() - control all non-data display. Use theme() if you just need to tweak the display of an existing, we have theme.minimal,classic,bw,light etc

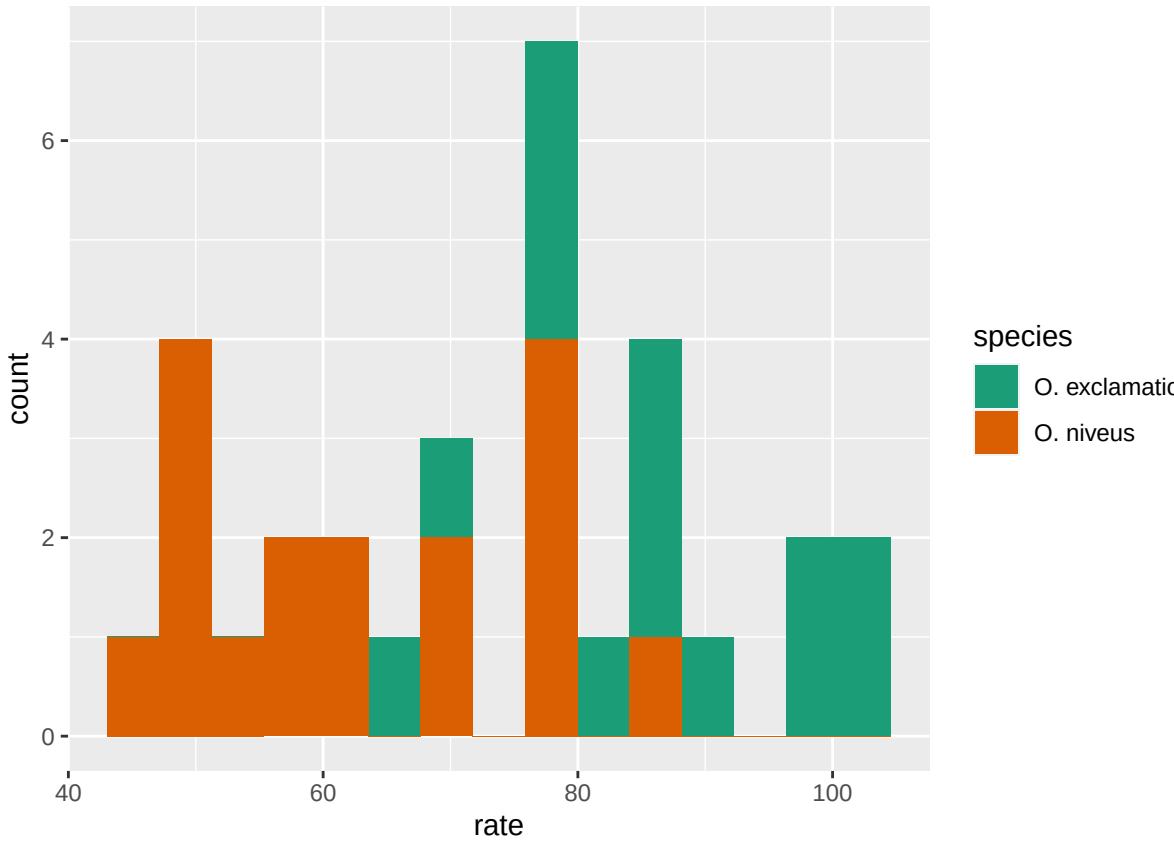
```
ggplot(crickets, aes(x = species,
                     y = rate,
                     color = species)) +
  geom_boxplot(show.legend = FALSE) +
  scale_color_brewer(palette = "Dark2") +
  theme_minimal()
```



To remove the gray background in the chart using theme minimal

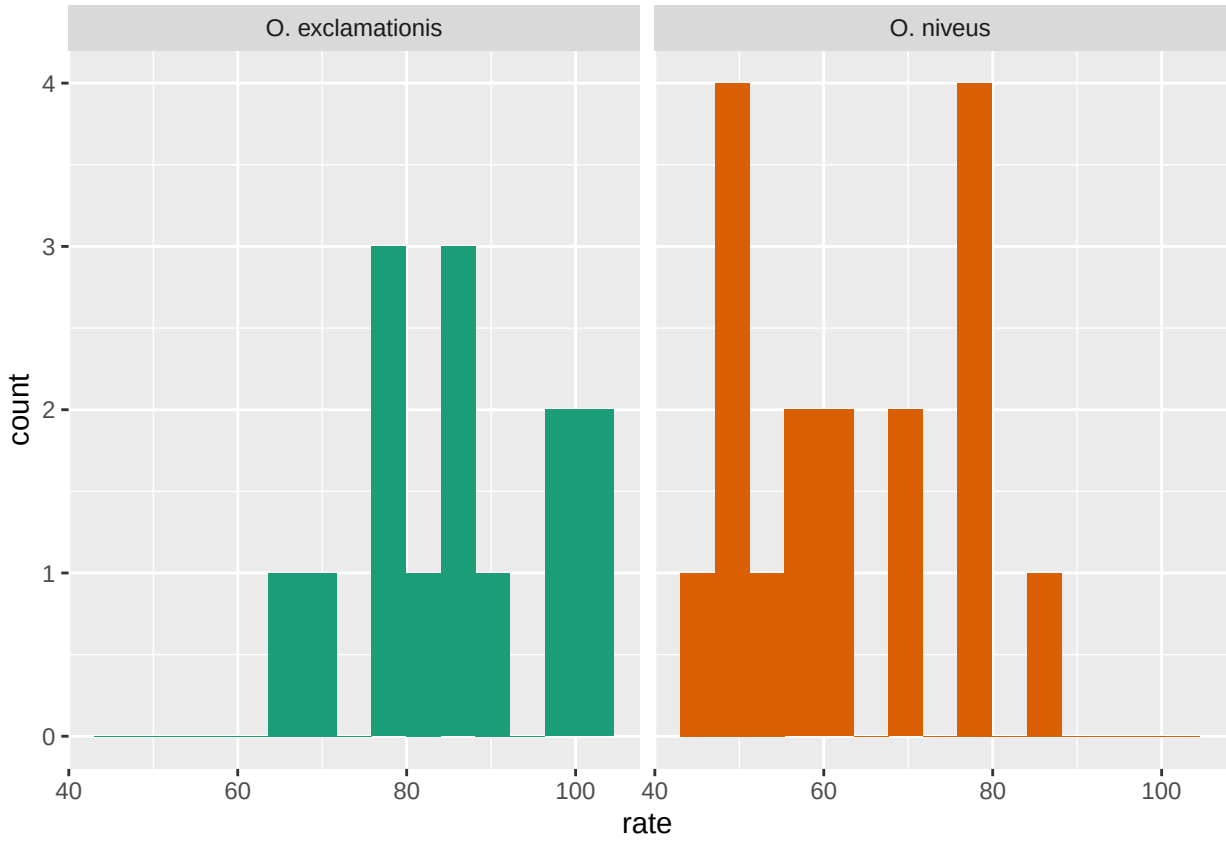
Faceting

```
ggplot(crickets, aes(x = rate,
                     fill = species)) +
  geom_histogram(bins = 15) +
  scale_fill_brewer(palette = "Dark2")
```



Without Faceting

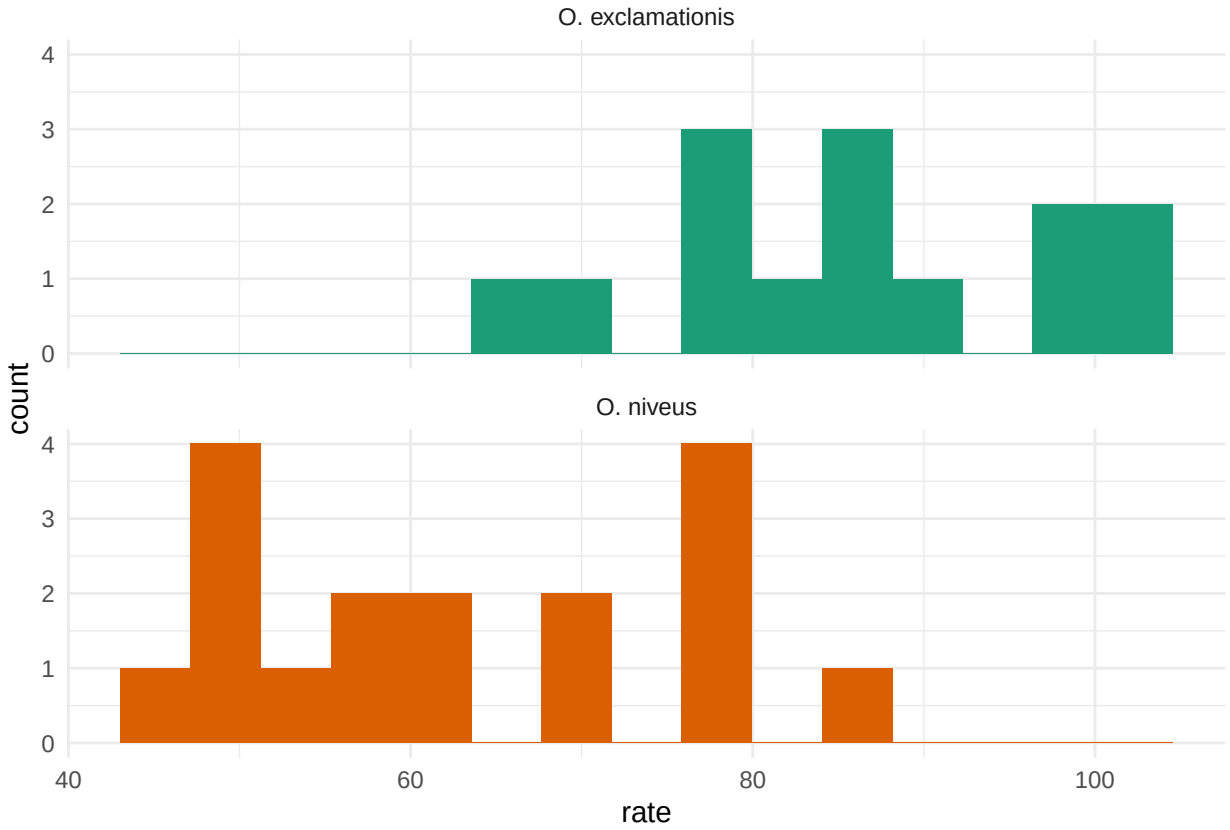
```
ggplot(crickets, aes(x = rate,
                     fill = species)) +
  geom_histogram(bins = 15,
                 show.legend = FALSE) +
  facet_wrap(~species) +
  scale_fill_brewer(palette = "Dark2")
```



With faceting

To have just one column - the second bar will be below the first

```
ggplot(crickets, aes(x = rate,
                     fill = species)) +
  geom_histogram(bins = 15,
                 show.legend = FALSE) +
  facet_wrap(~species,
             ncol = 1) +
  scale_fill_brewer(palette = "Dark2") +
  theme_minimal()
```


O. exclamationis
O. niveus
count
rate
0
1
2
3
4
40
60
80
100