

# Continuous data

---

`geom_histogram()`

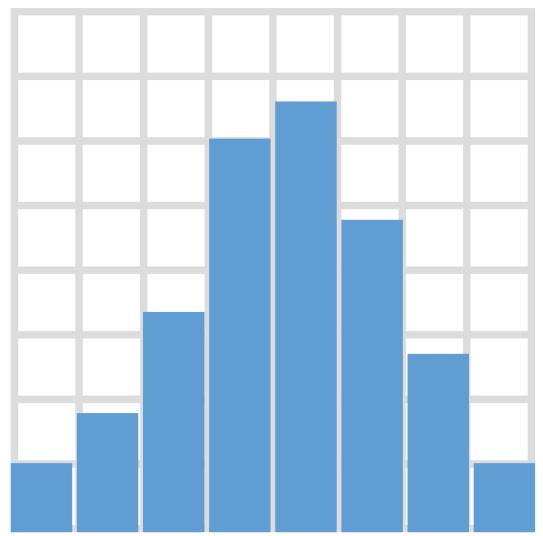
`geom_density()`

`geom_boxplot()`

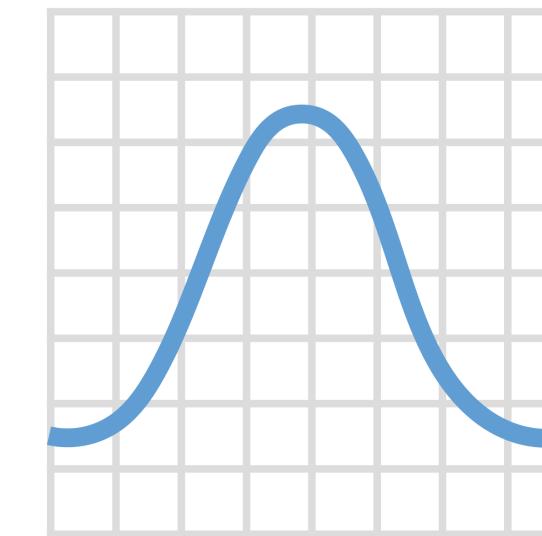
`geom_point()`

# GEOMS for continuous data

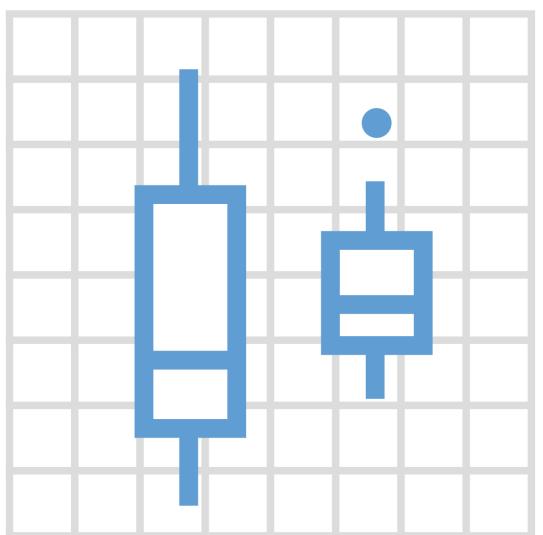
---



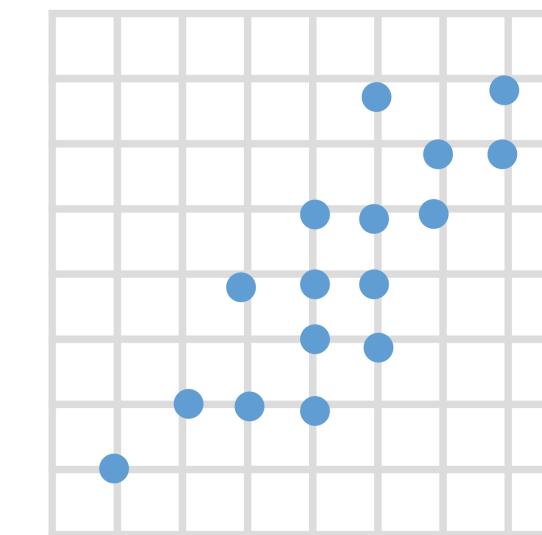
`geom_histogram()`



`geom_density()`



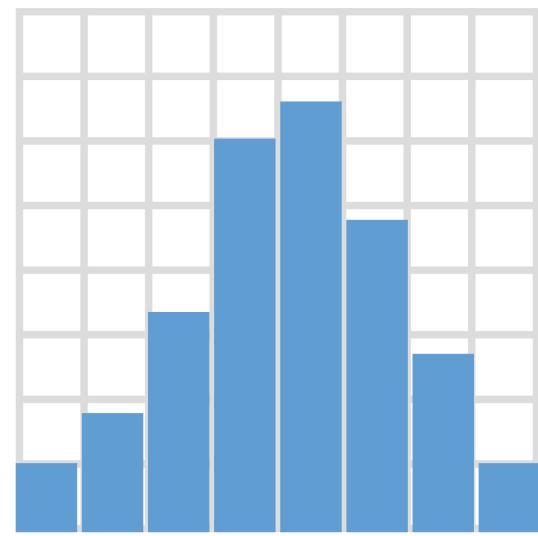
`geom_boxplot()`



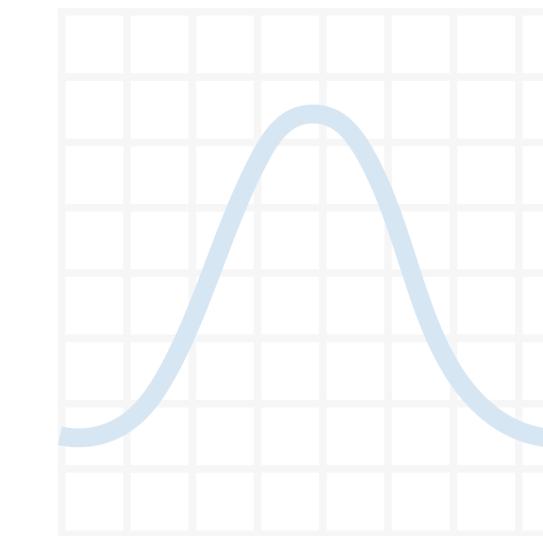
`geom_point()`

# GEOMS for continuous data

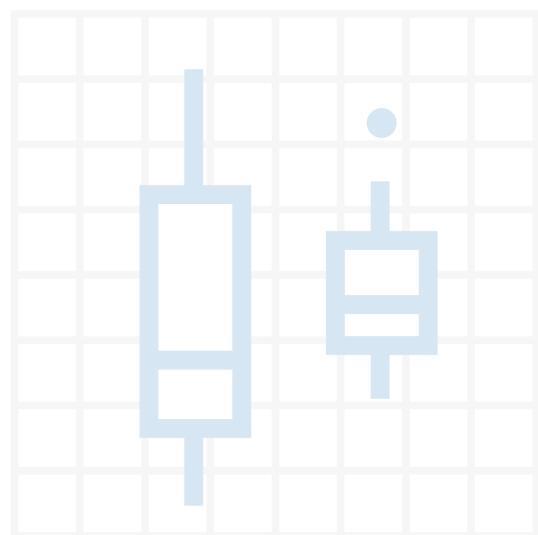
---



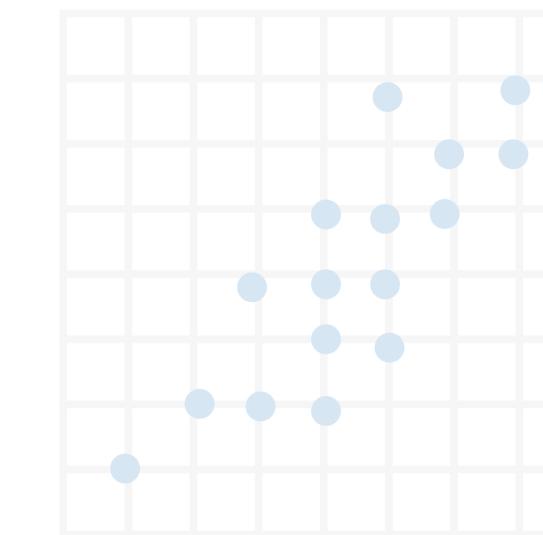
`geom_histogram()`



`geom_density()`

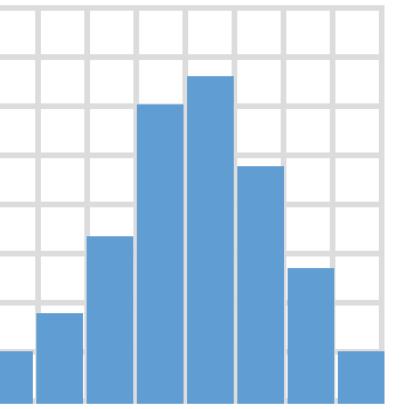


`geom_boxplot()`

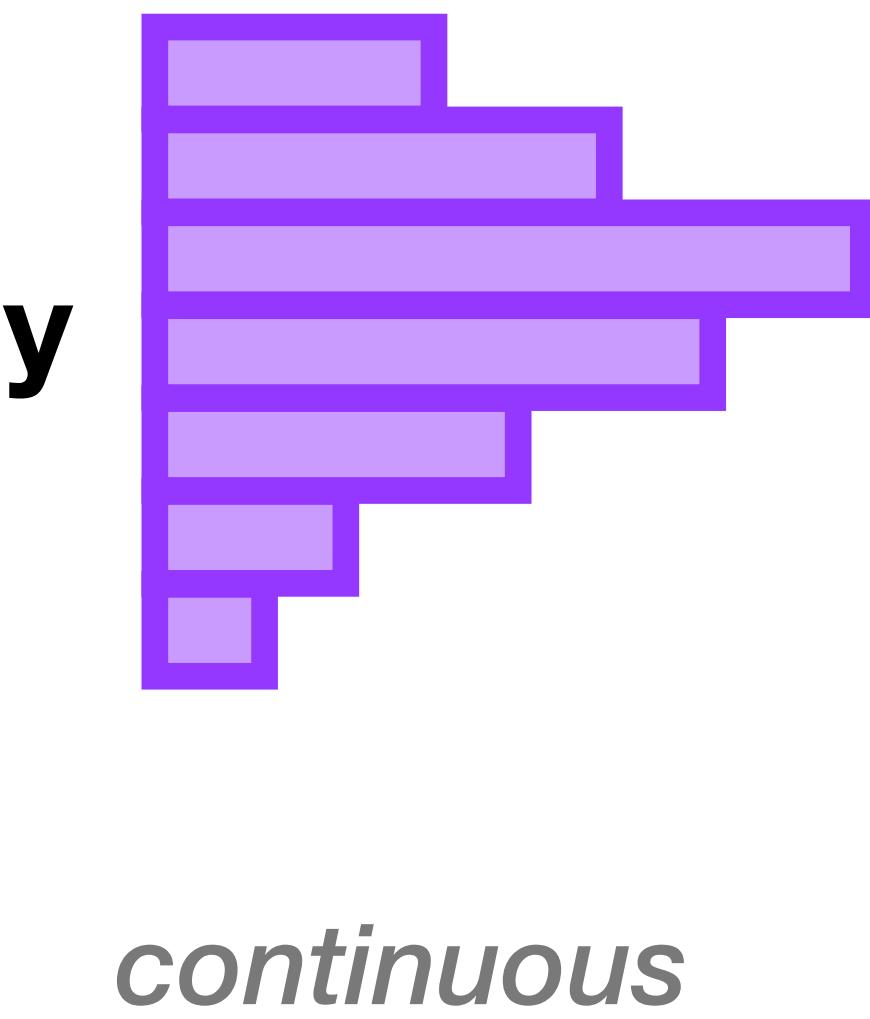
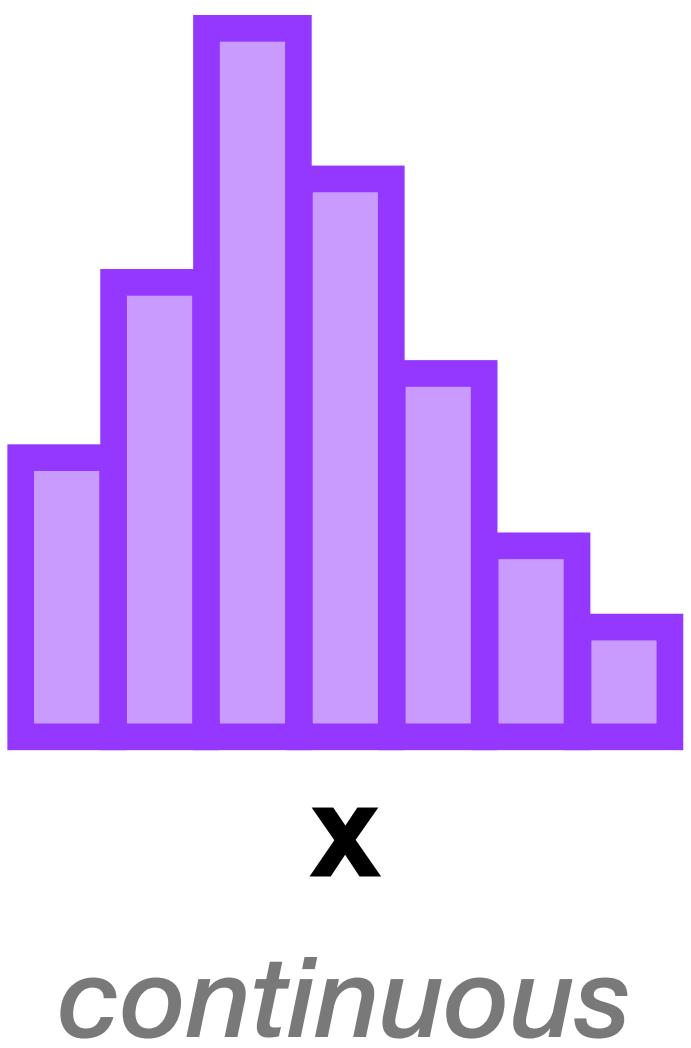


`geom_point()`

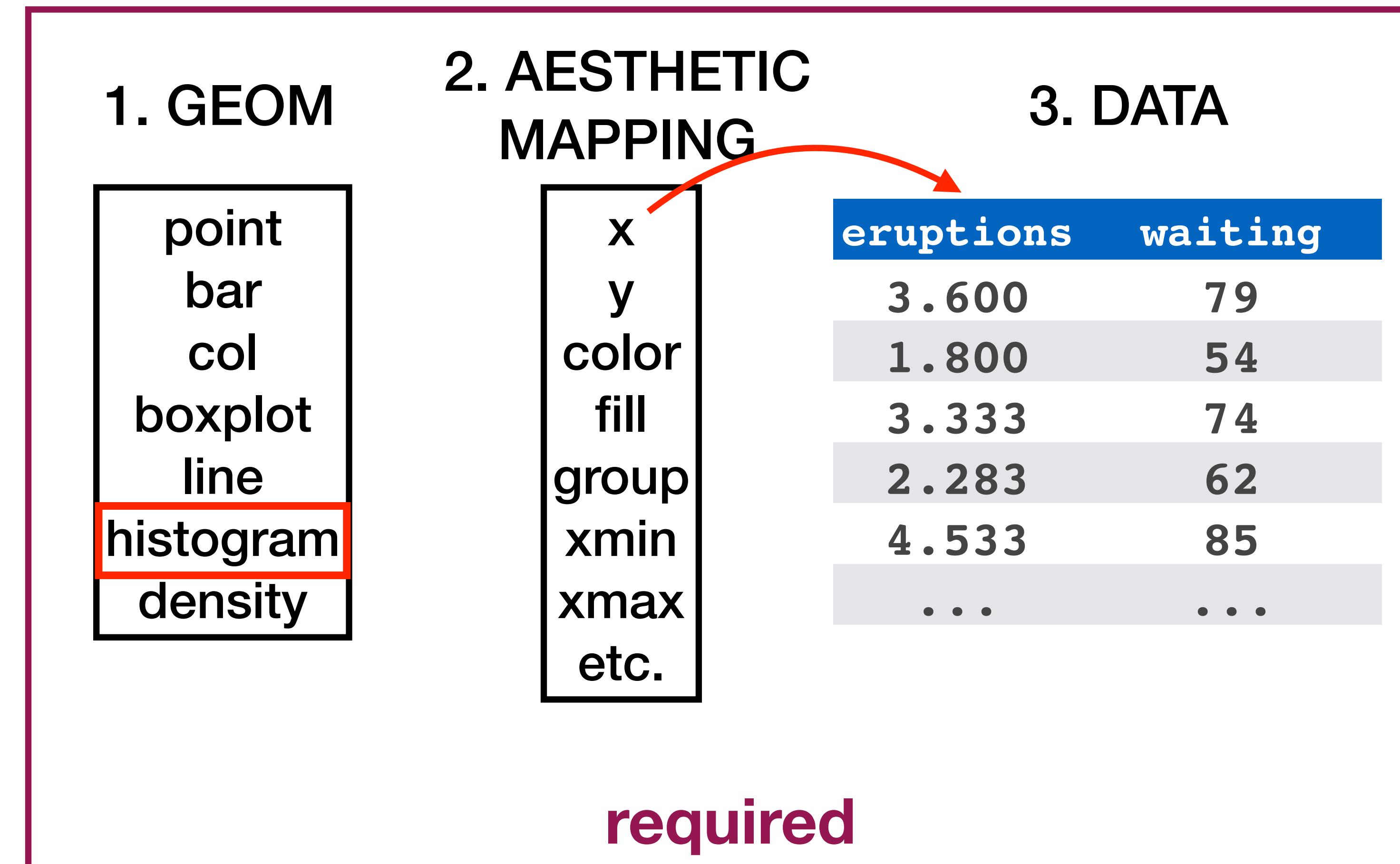
# geom\_histogram()

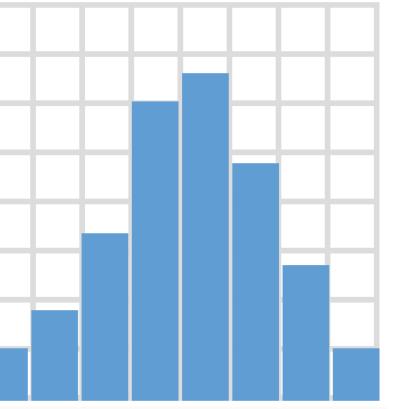


- Shows the distribution of a continuous variable (unbinned = no count column)
- Requires an **x** (vertical bars) or **y** (horizontal bars, rare)
- No spaces between bars



# Putting it all together: start with the GEOM





# Look at the data

```
str(faithful)
#> 'data.frame': 272 obs. of 2 variables:
#> $ eruptions: num 3.6 1.8 3.33 2.28 4.53 ...
#> $ waiting   : num 79 54 74 62 85 55 88 85 51 85
...
...
```

Remember: data must be continuous (numeric)!

# Ready to code

```
initialize          plot      data frame      aesthetic mapping  
↓                  ↓                      ↓  
ggplot(faithful, aes(x = eruptions)) +  
geom_histogram()  
↑  
geom
```

The geom *inherits* data and mappings  
from the call to `ggplot()`

# Do not start a new line with "+"



Pitfall  
Alert!

*initialize  
plot*

*data frame*

*aesthetic mapping*

```
ggplot(faithful, aes(x = eruptions)) +  
  geom_histogram()
```

*geom*

Good practice: end lines with "+"

# If you do...



```
ggplot(faithful, aes(x = eruptions))  
+ geom_histogram()
```

*Error* in `+.gg`:  
! Cannot use `+` with a single argument  
i Did you accidentally put `+` on a new line?  
Run `rlang::last\_trace()` to see where the error occurred.

# With parameter names

*initialize*

*plot*

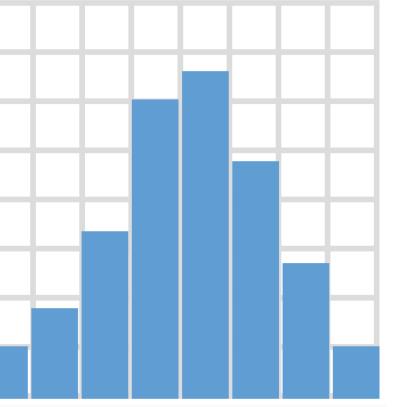
*data frame*

*aesthetic mapping*

```
ggplot(data = faithful, mapping = aes(x = eruptions)) +  
  geom_histogram()
```

*geom*

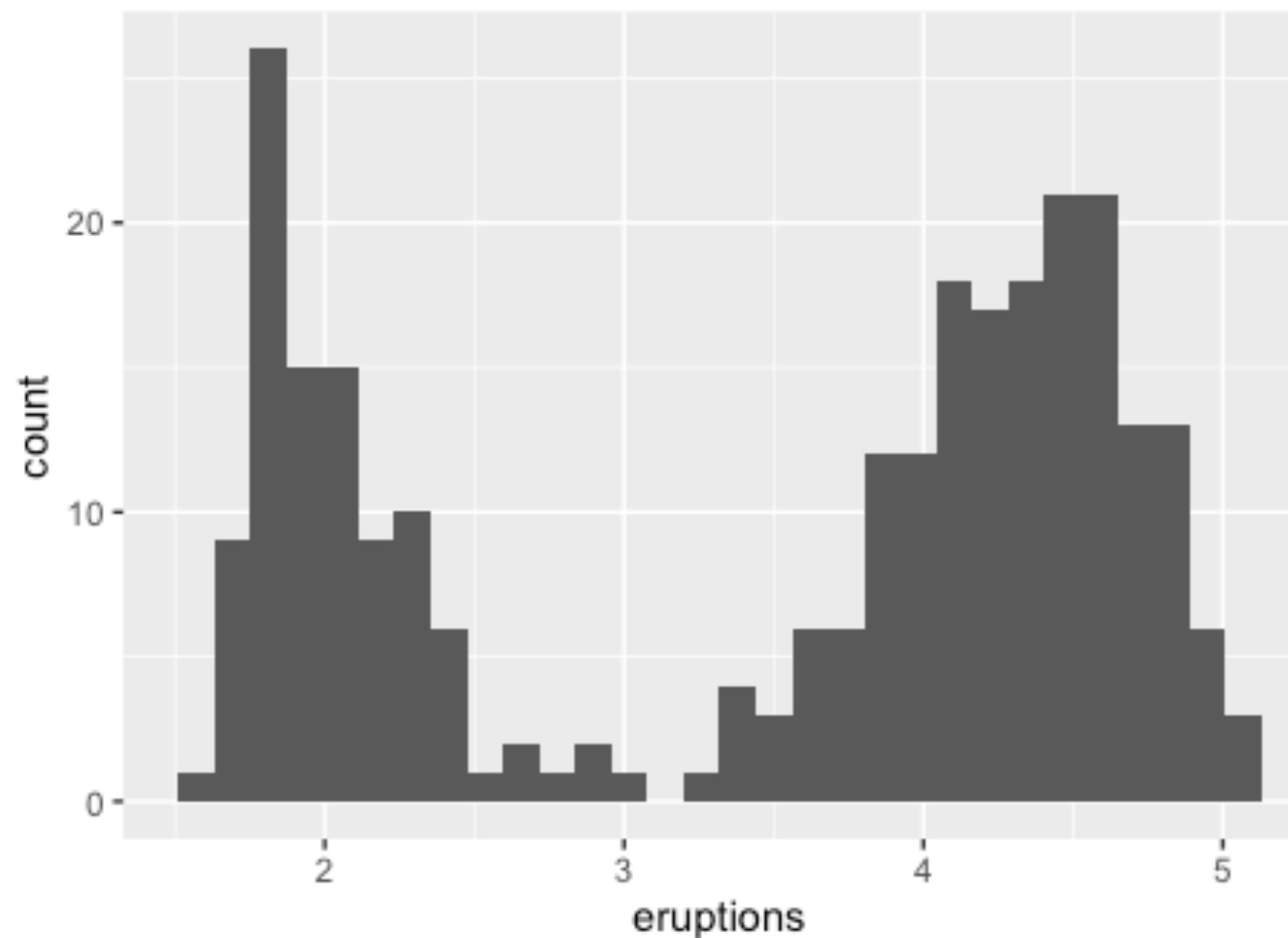
Not generally used EXCEPT  
for "data =" in GEOMS --  
will discuss later



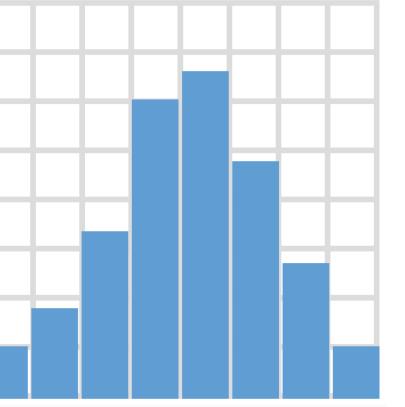
# geom\_histogram()

```
ggplot(faithful, aes(x = eruptions)) +  
  geom_histogram()  
#> `stat_bin()` using `bins = 30`. Pick better value with  
`binwidth`.
```

Let's change  
the color  
and fill

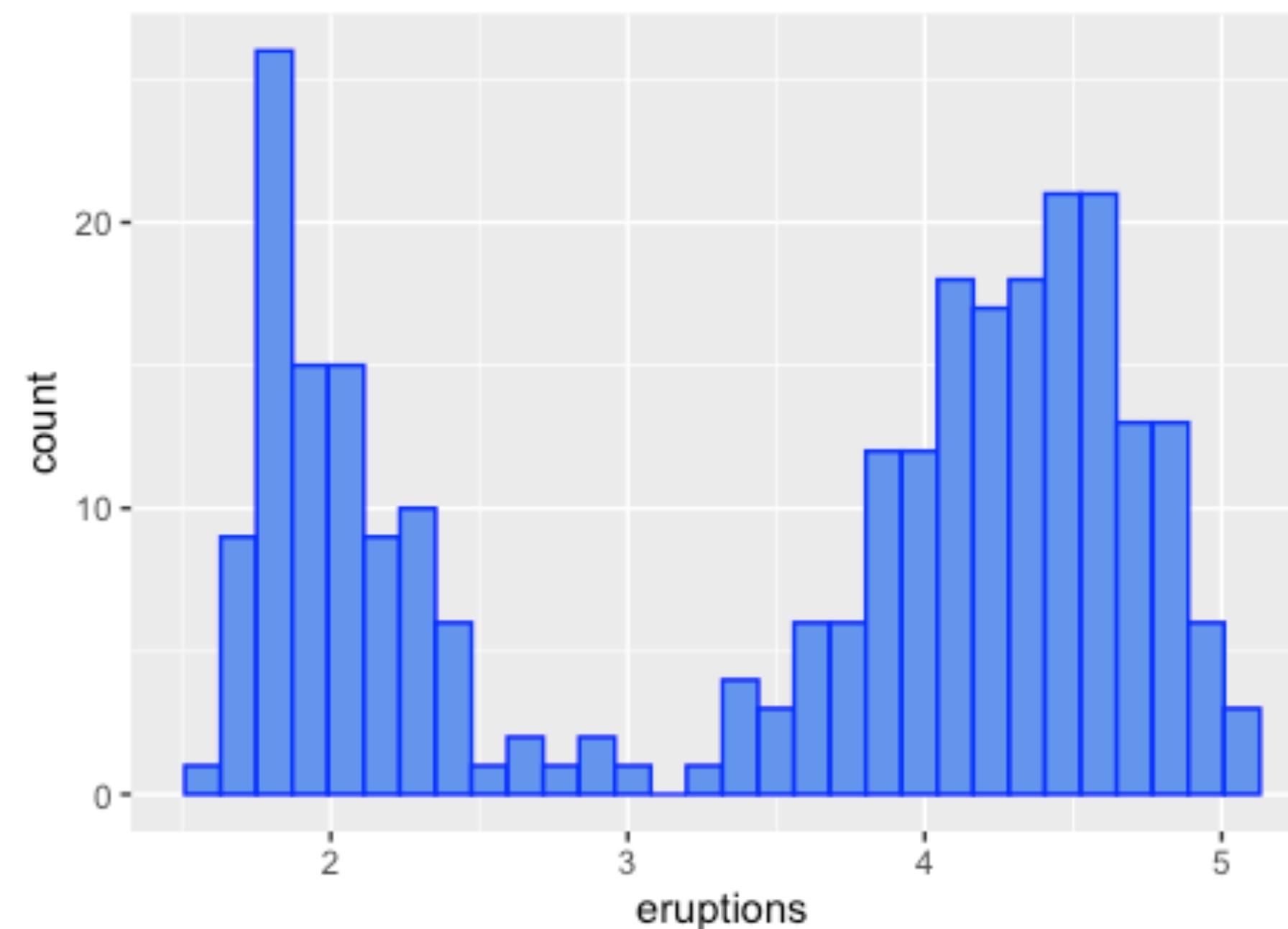


hist() ggplot2  
border color  
col fill



# Change the color and fill

```
ggplot(faithful, aes(x = eruptions)) +  
  geom_histogram(color = "blue", fill = "cornflowerblue")  
#> `stat_bin()` using `bins = 30`. Pick better value with  
`binwidth`.
```

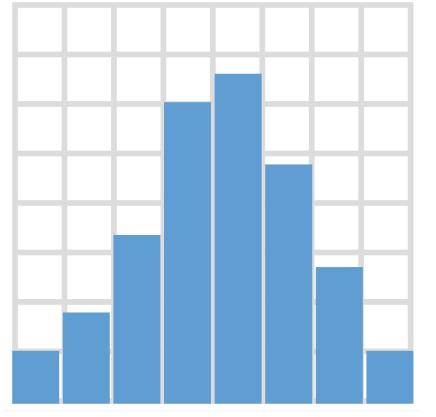


# Settings vs. mappings



```
ggplot(faithful, aes(x = eruptions)) +  
  geom_histogram(color = "blue", fill = "cornflowerblue")
```

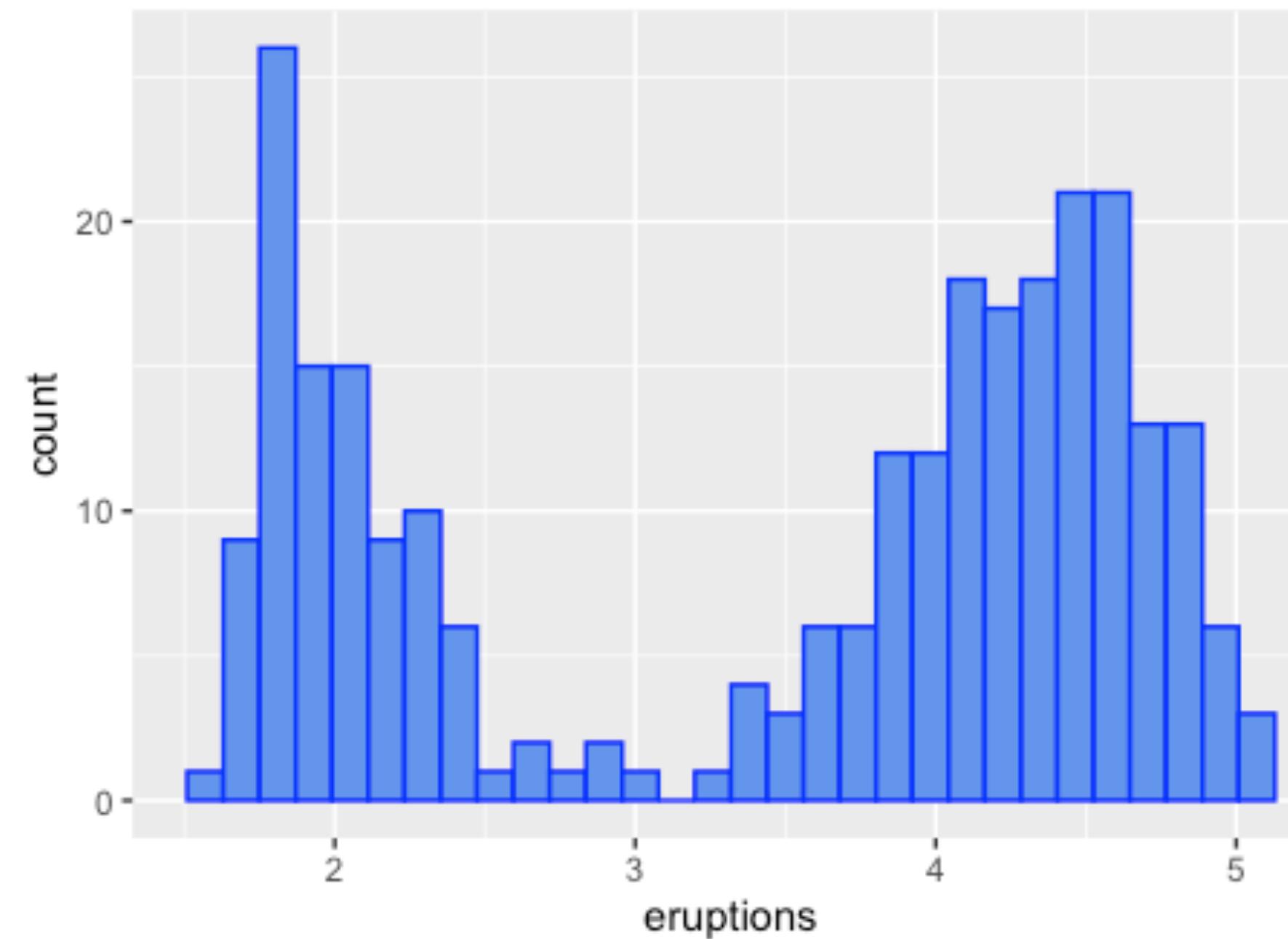
In this case, `color` and `fill` are *settings* -- think constants. They are *not* aesthetic mappings.

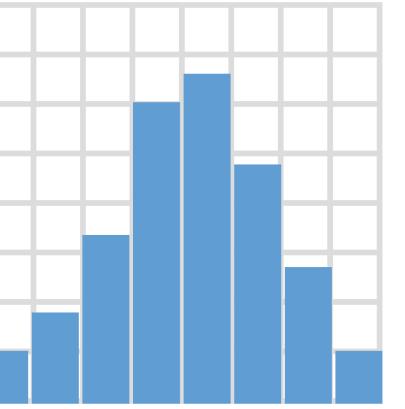


# geom\_histogram()

```
ggplot(faithful, aes(x = eruptions)) +  
  geom_histogram(color = "blue", fill = "cornflowerblue")  
#> `stat_bin()` using `bins = 30`. Pick better value with  
`binwidth`.
```

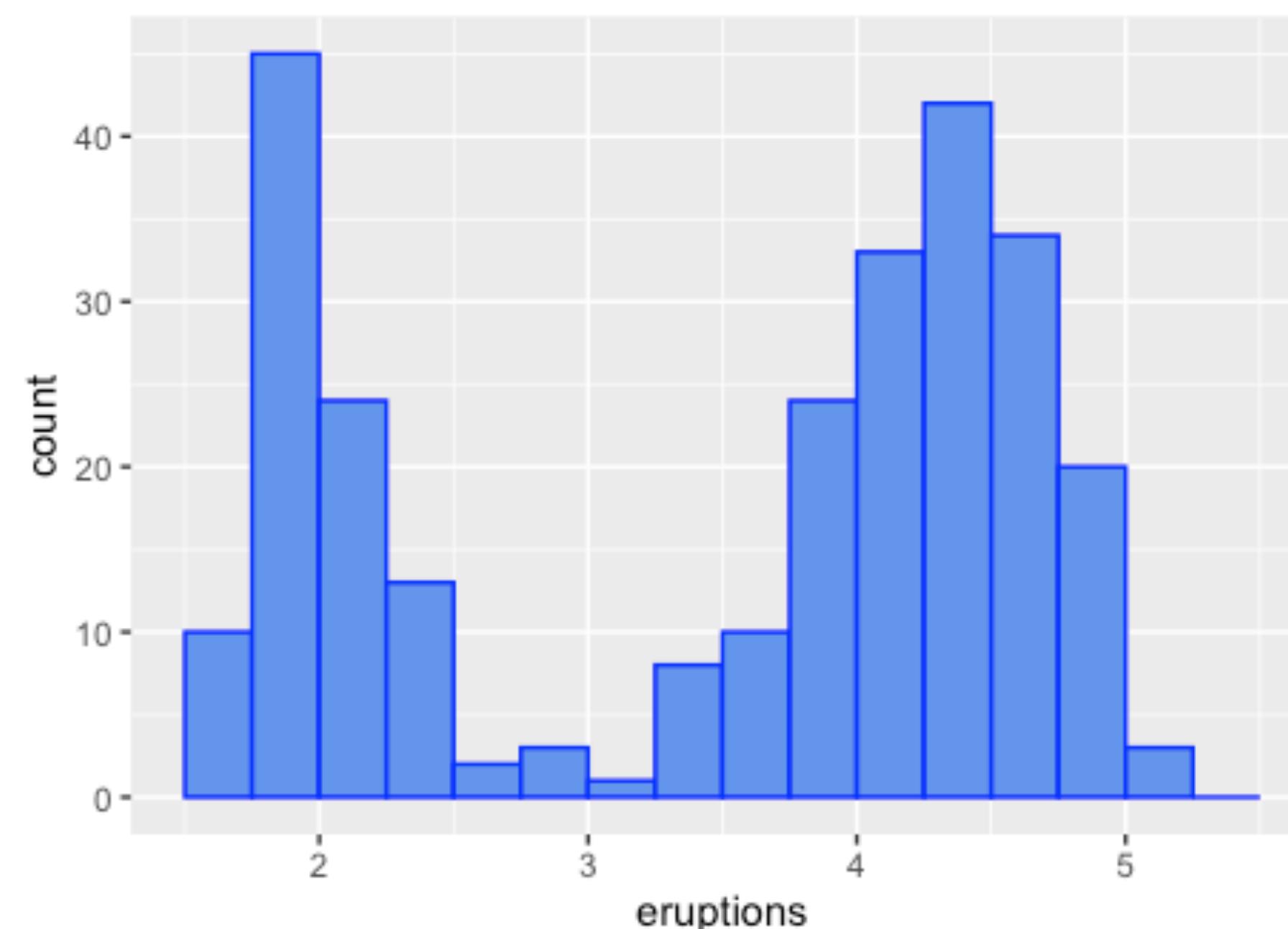
Let's vary  
the bin  
boundaries

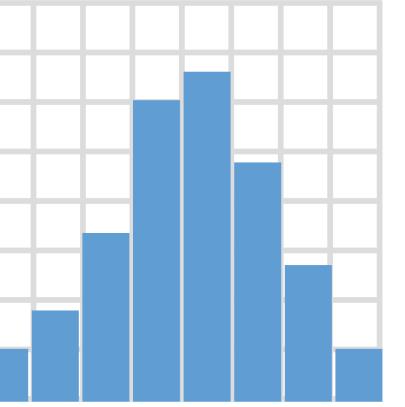




# New bin boundaries

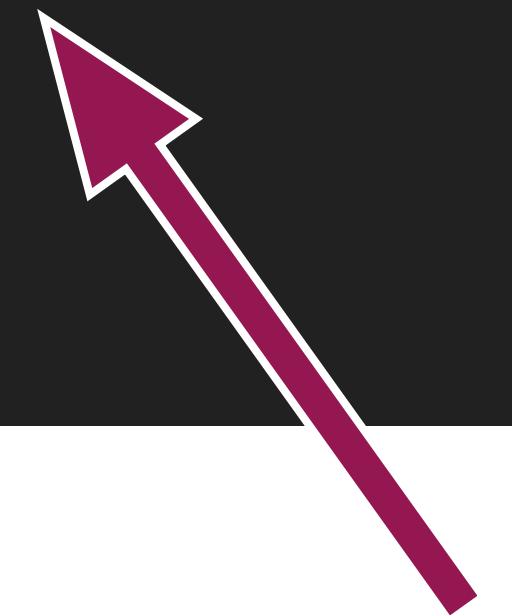
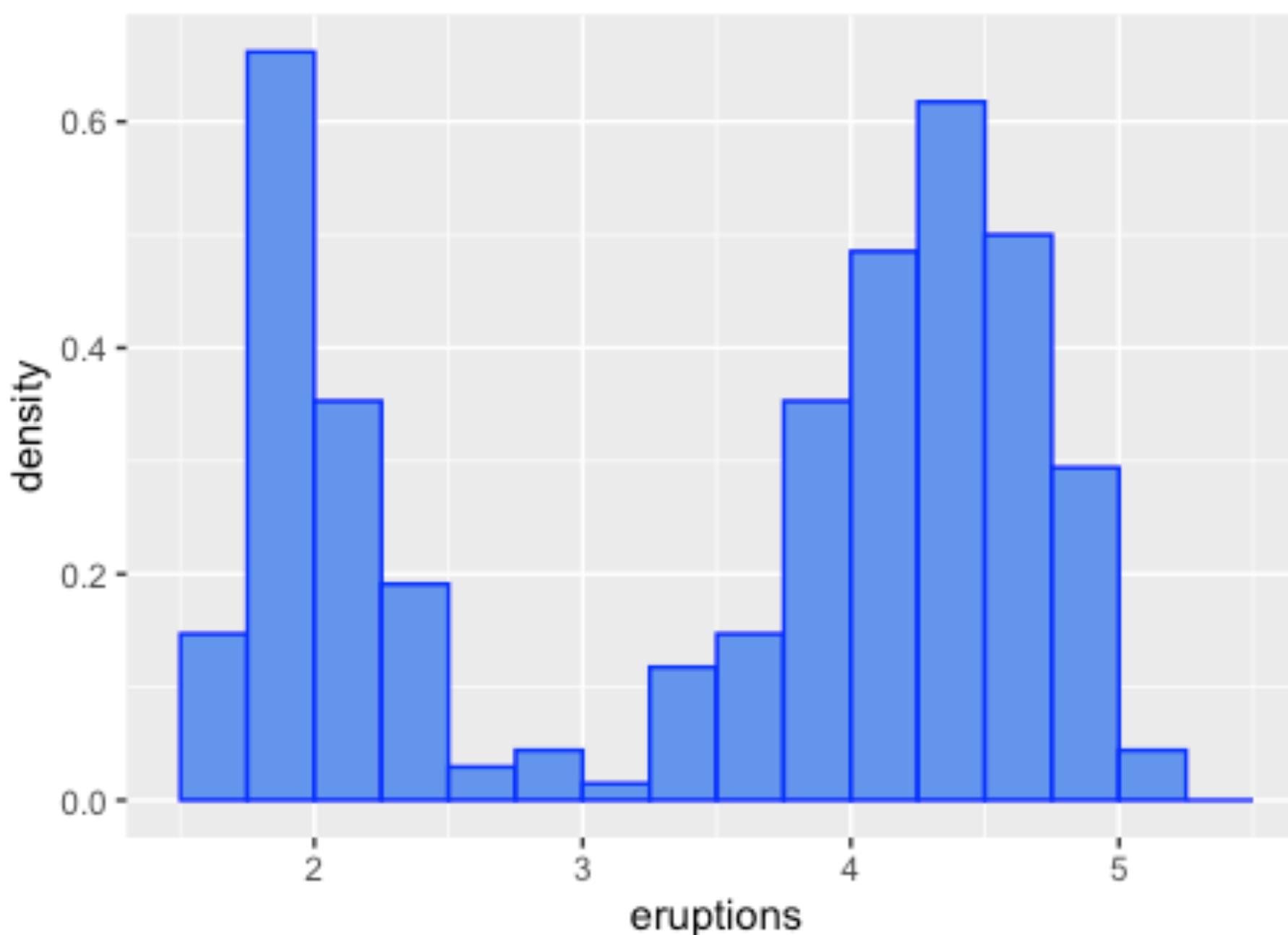
```
ggplot(faithful, aes(x = eruptions)) +  
  geom_histogram(breaks = seq(1.5, 5.5, .25),  
                 color = "blue",  
                 fill = "cornflowerblue")
```





# Density histogram

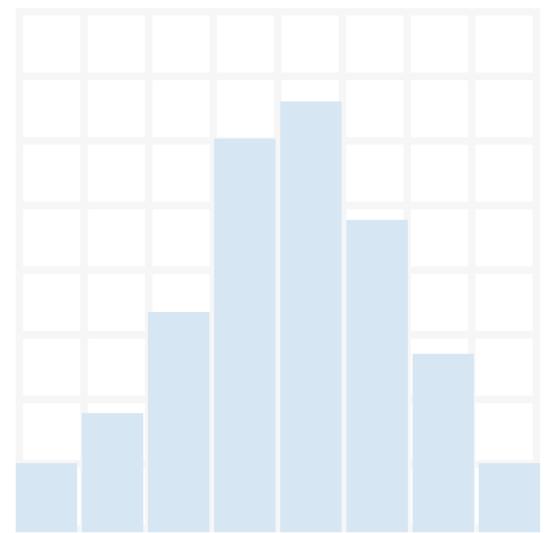
```
ggplot(faithful, aes(x = eruptions, y = after_stat(density))) +  
  geom_histogram(breaks = seq(1.5, 5.5, .25),  
                 color = "blue",  
                 fill = "cornflowerblue")
```



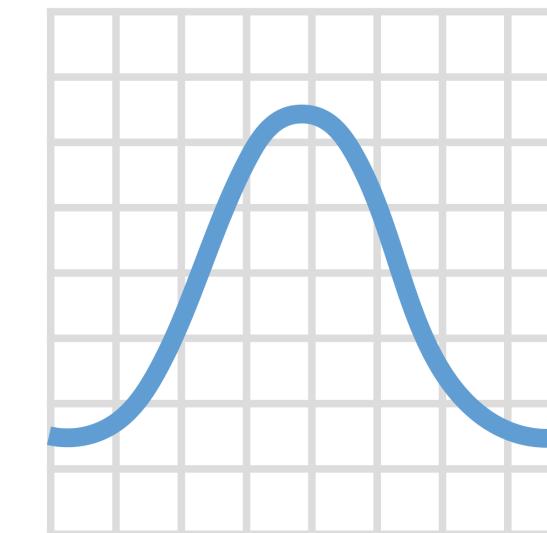
Replaces the default:  
`after_stat(count)`  
  
(not very common  
except for histograms)

# GEOMS for continuous data

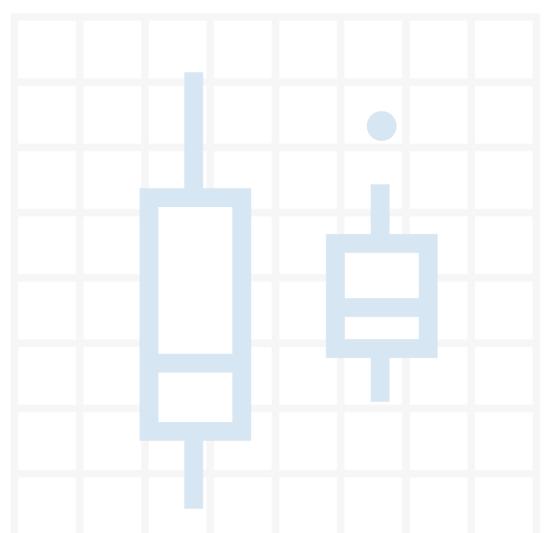
---



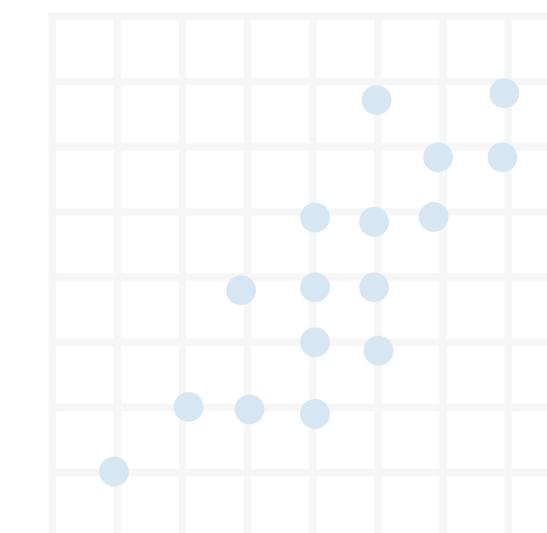
`geom_histogram()`



`geom_density()`

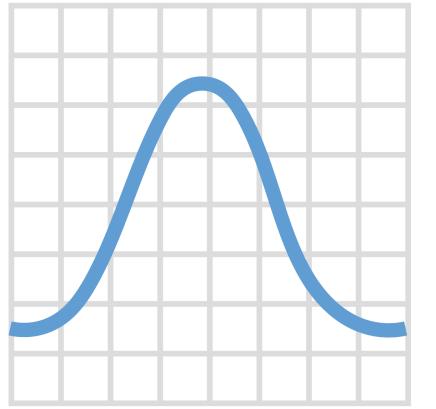


`geom_boxplot()`

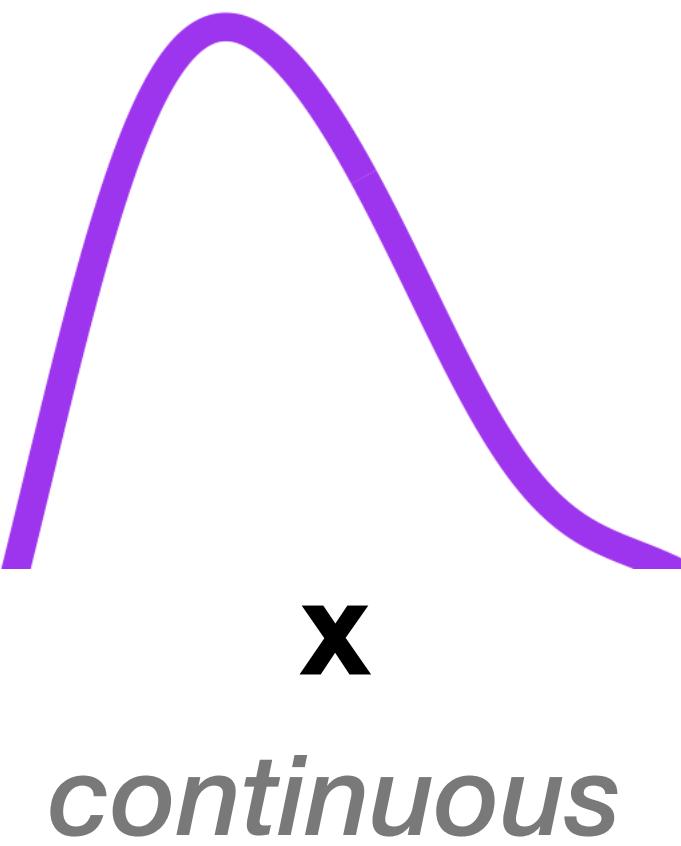


`geom_point()`

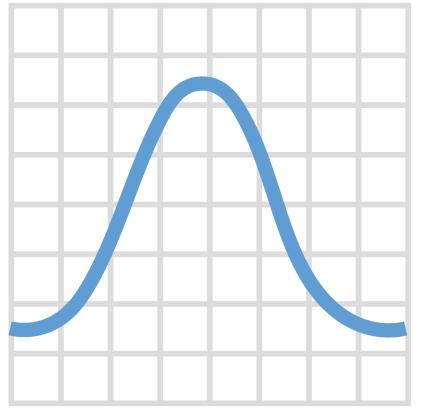
# geom\_density()



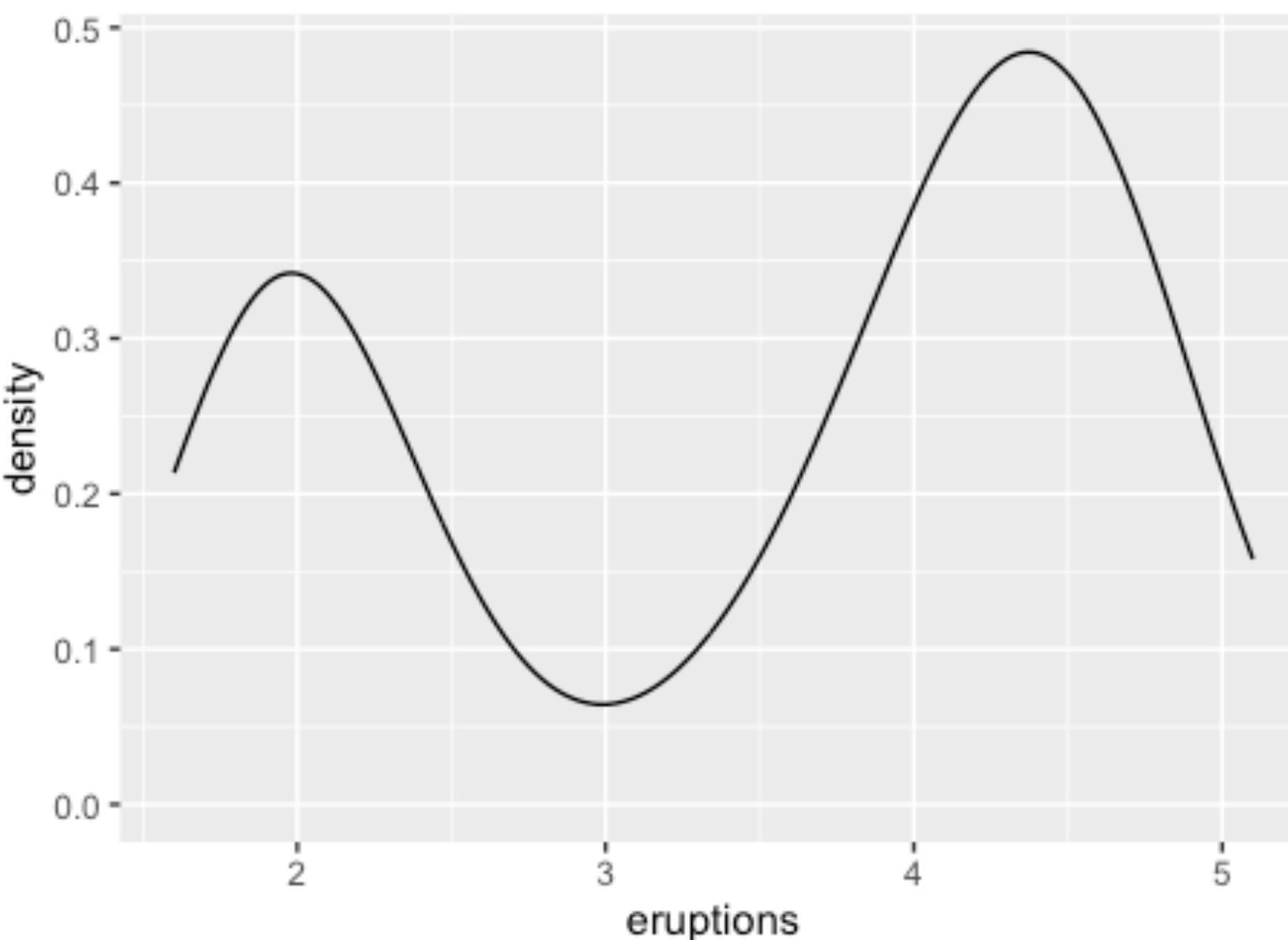
- Like a histogram, shows the distribution of a continuous variable
- Requires an **x** or **y** (rare) mapping
- The other axis (usually **y**) defaults to a density scale



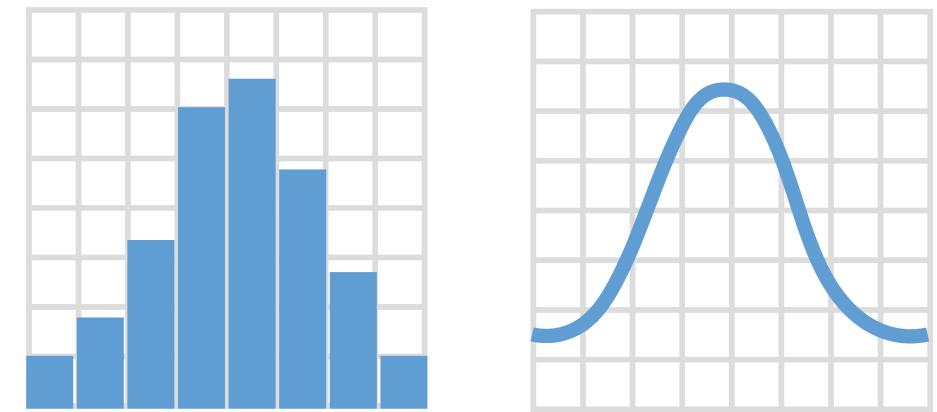
# geom\_density()



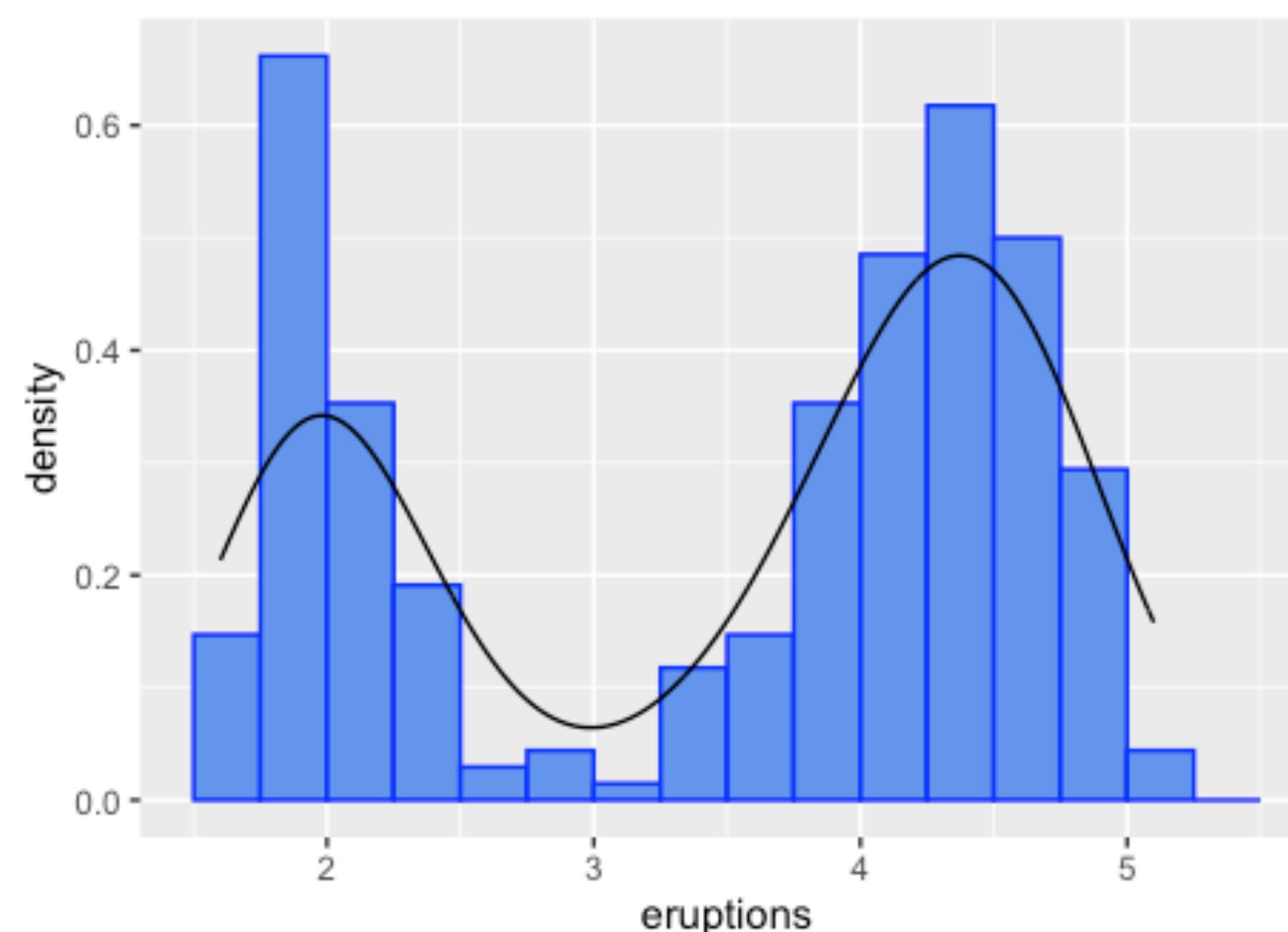
```
ggplot(faithful, aes(x = eruptions)) +  
  geom_density()
```



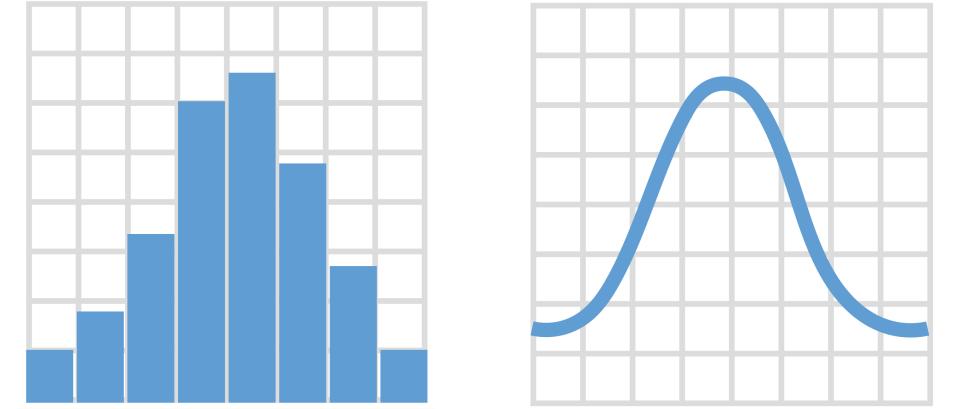
# Two GEOMs



```
ggplot(faithful, aes(x = eruptions, y = after_stat(density)) +  
  geom_histogram(breaks = seq(1.5, 5.5, .25), color = "blue",  
                 fill = "cornflowerblue") +  
  geom_density()
```

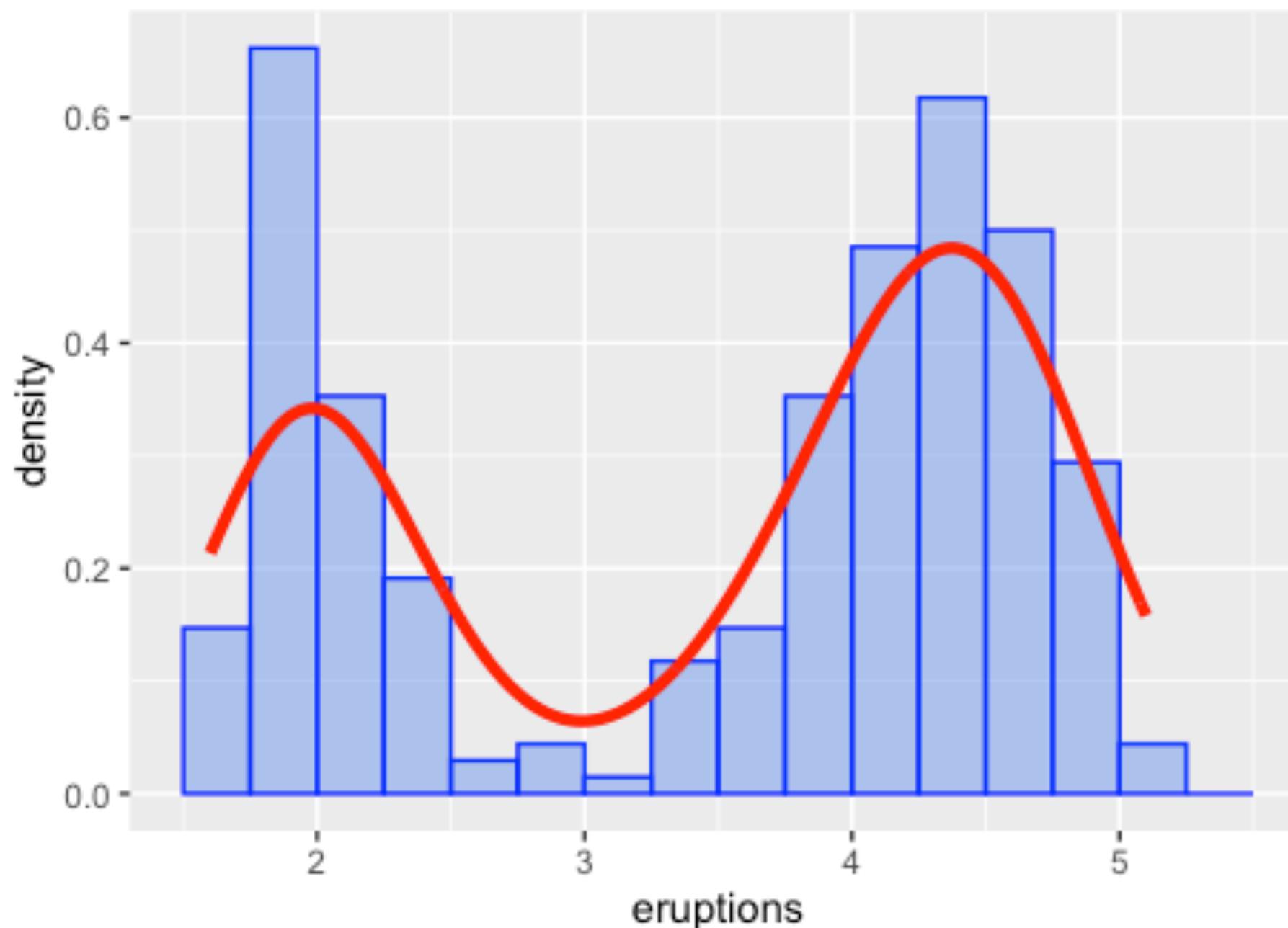


# Set linewidth and color

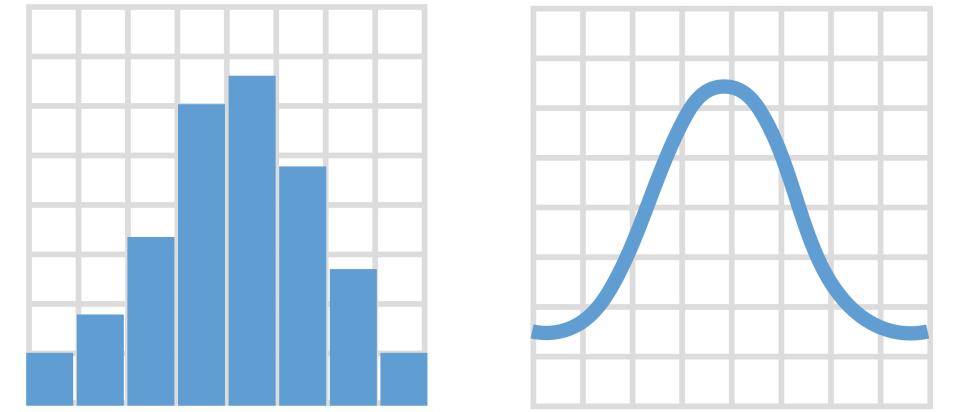


```
ggplot(faithful, aes(x = eruptions, y = after_stat(density))) +  
  geom_histogram(breaks = seq(1.5, 5.5, .25), color = "blue",  
                 fill = "cornflowerblue", alpha = .5) +  
  geom_density(linewidth = 1.5, color = "red")
```

- ▶ `lwd` also works for `linewidth`
- ▶ default `linewidth` is 0.5

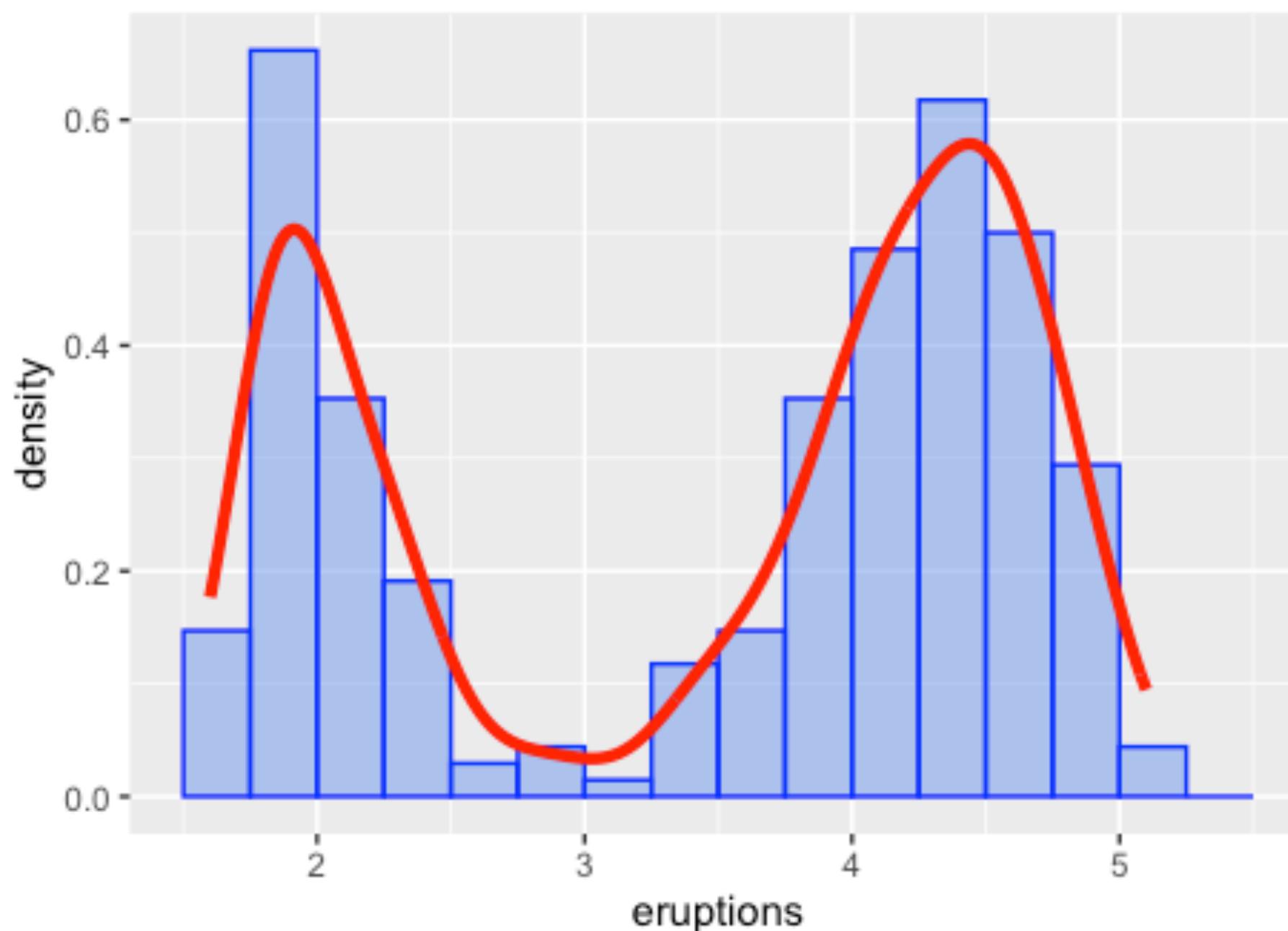


# Change the bandwidth

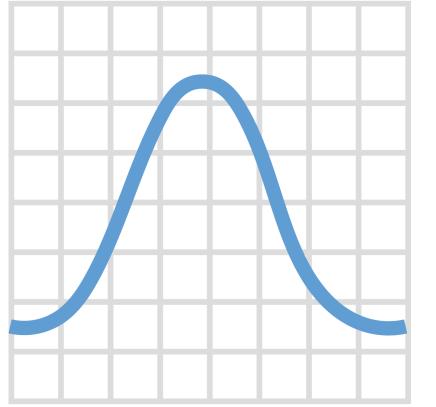


```
ggplot(faithful, aes(x = eruptions, y = after_stat(density))) +  
  geom_histogram(breaks = seq(1.5, 5.5, .25), color = "blue",  
                 fill = "cornflowerblue", alpha = .5) +  
  geom_density(adjust = 0.5, linewidth = 1.5, color = "red")
```

Change kernel  
density estimate  
bandwidth with bw  
(absolute) or adjust  
(relative to default)  
see ?density

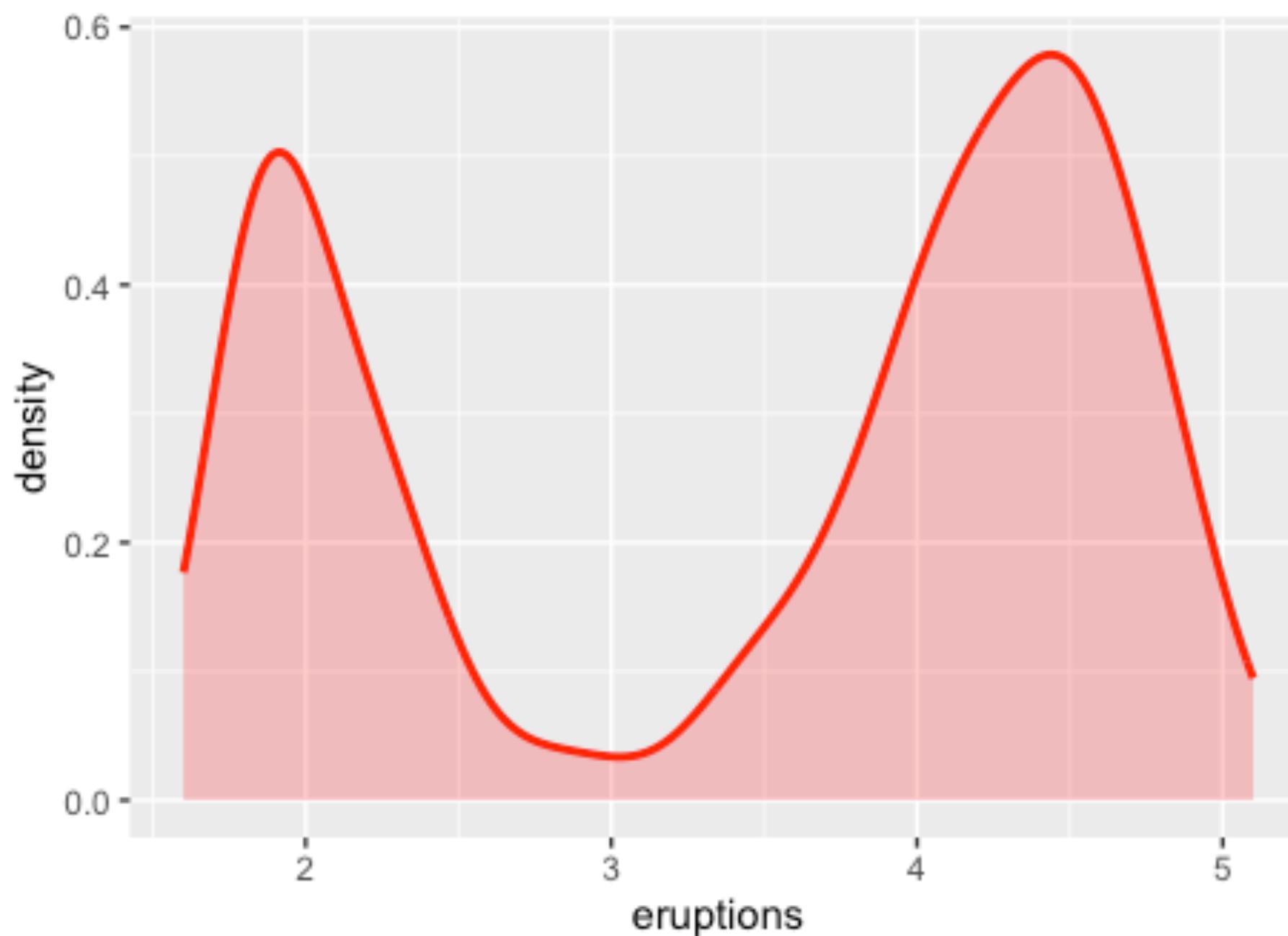


# Add a fill and alpha transparency



```
ggplot(faithful, aes(x = eruptions, y = after_stat(density))) +  
  geom_density(adjust = 0.5, linewidth = 1, color = "red",  
               fill = "red", alpha = .25)
```

The scale of alpha  
is 0 to 1 (0% to  
100% transparency)



# EXERCISES

---

- Code: [www.github.com/jtr13/SDSS2023](https://www.github.com/jtr13/SDSS2023)
- Open `geom_histogram.Rmd`
- Run the code.
- Make changes and see what happens.
- Try the exercises.
- Repeat with `geom_density.Rmd`