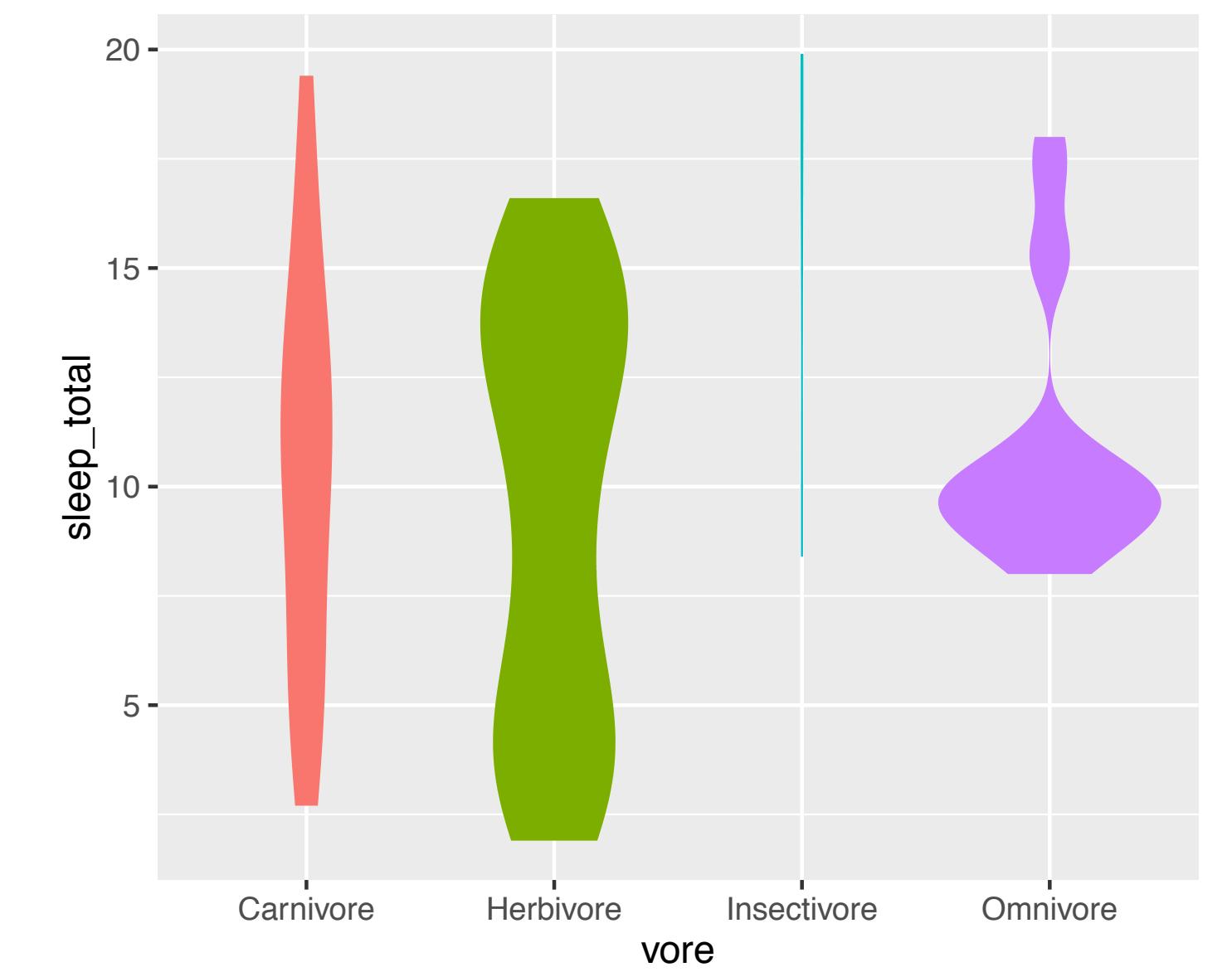
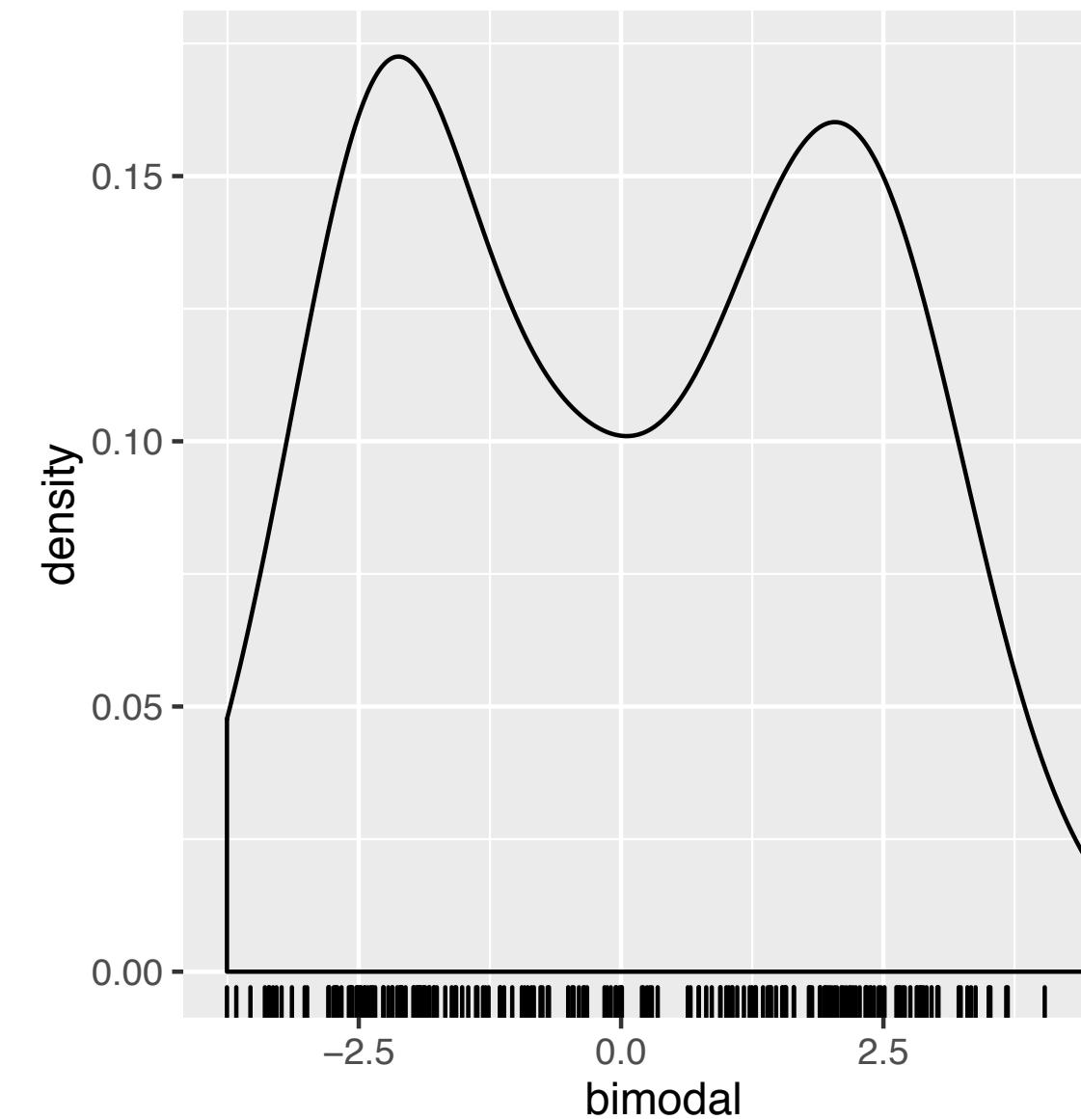
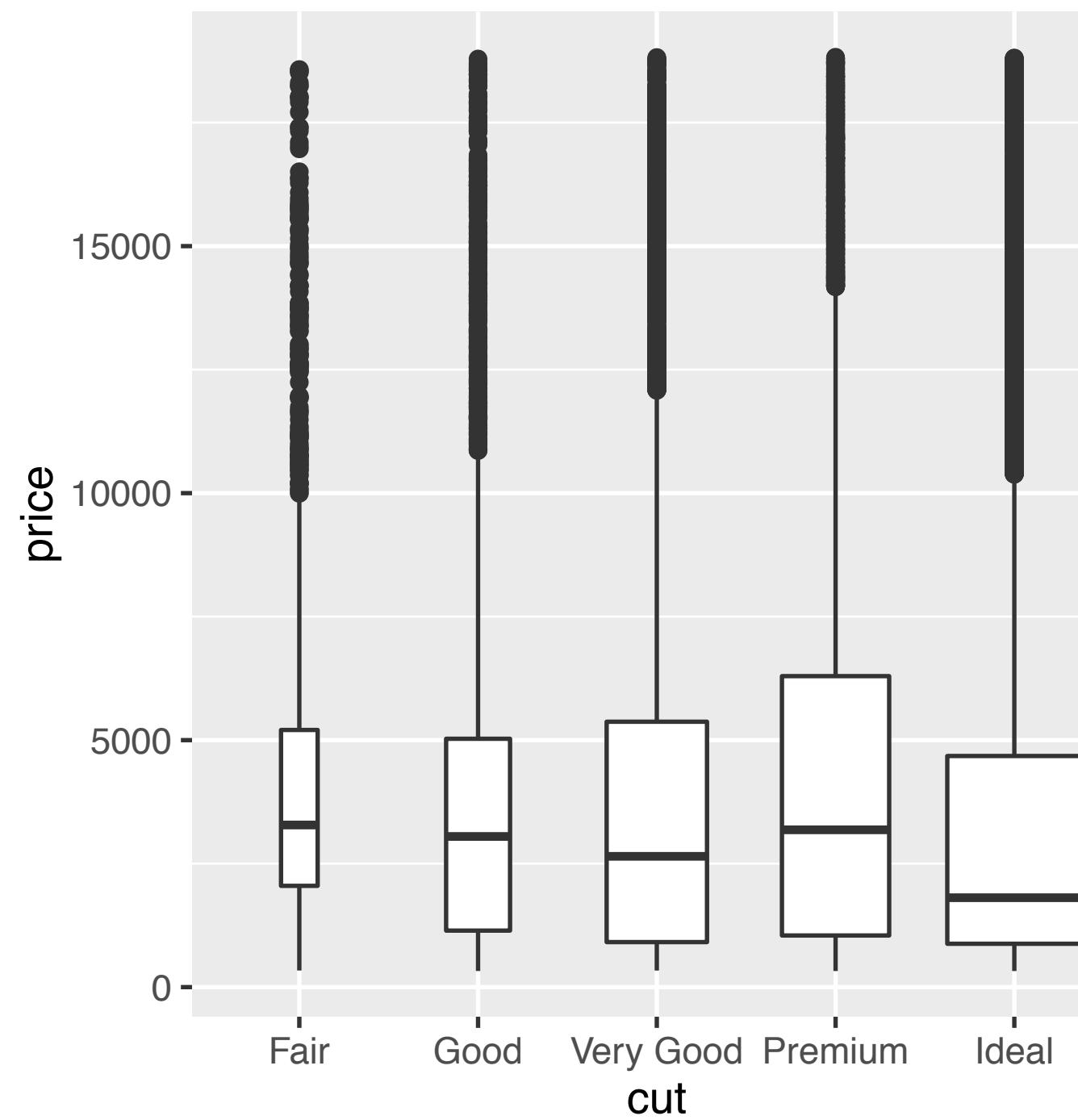




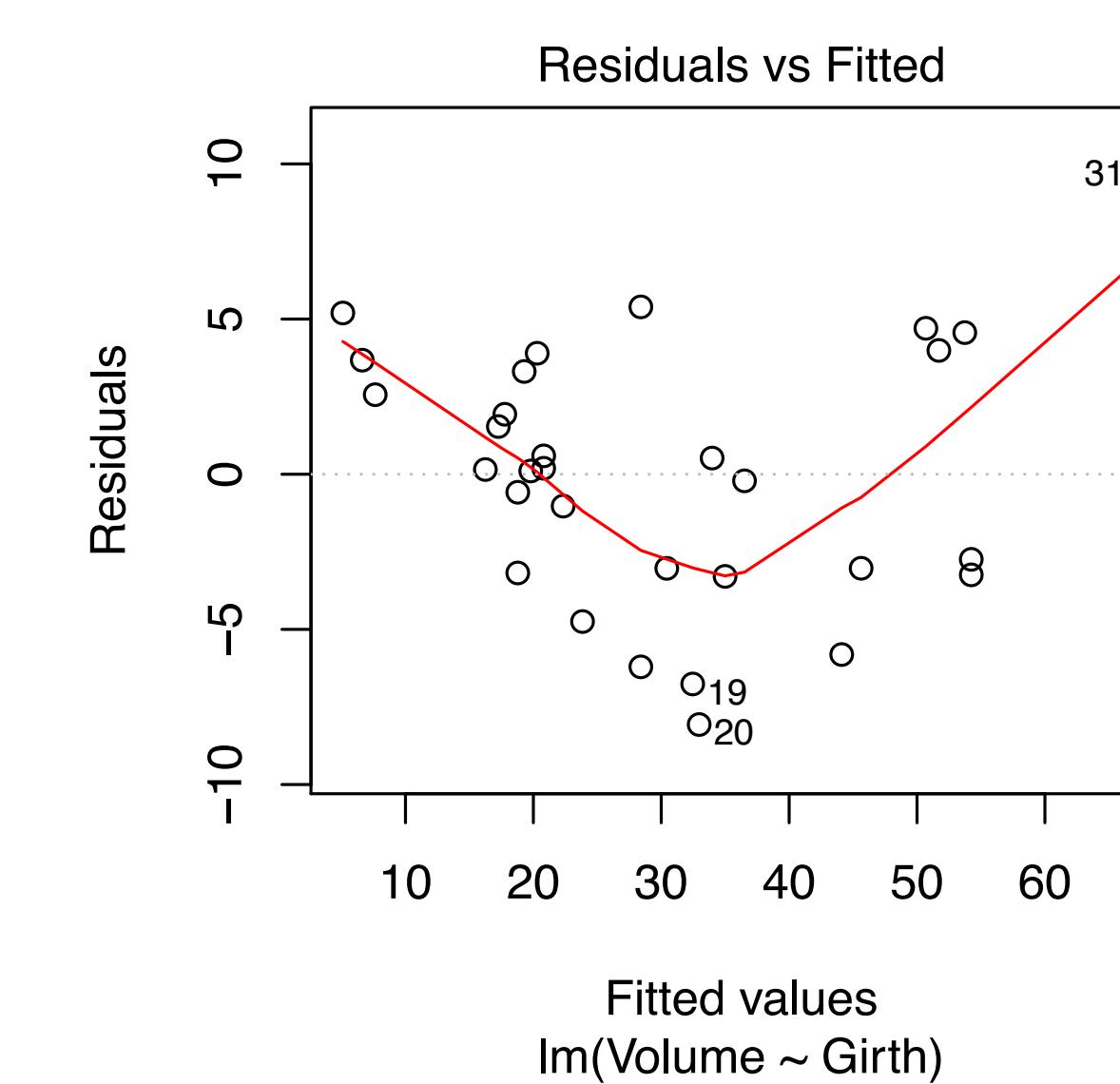
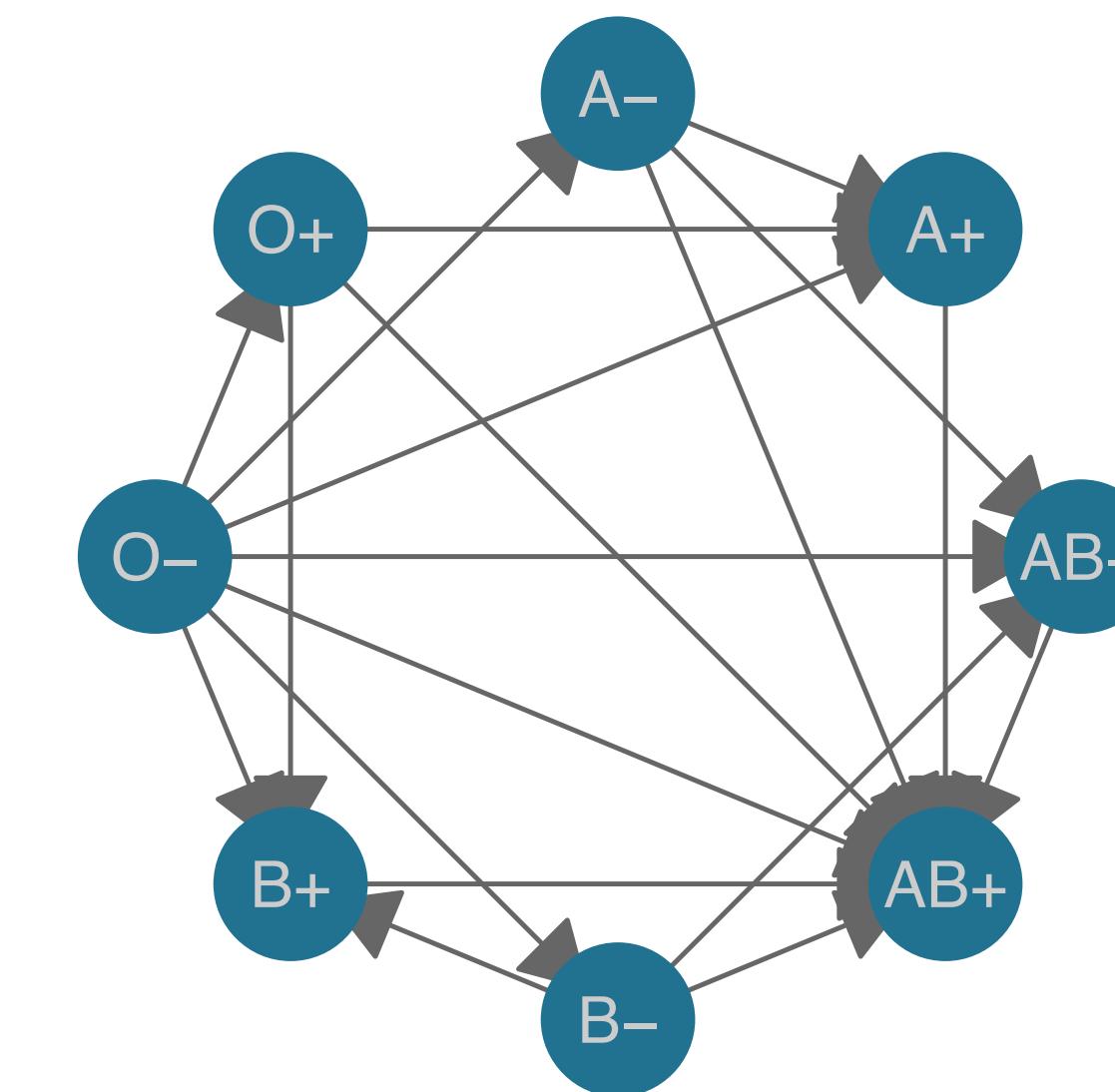
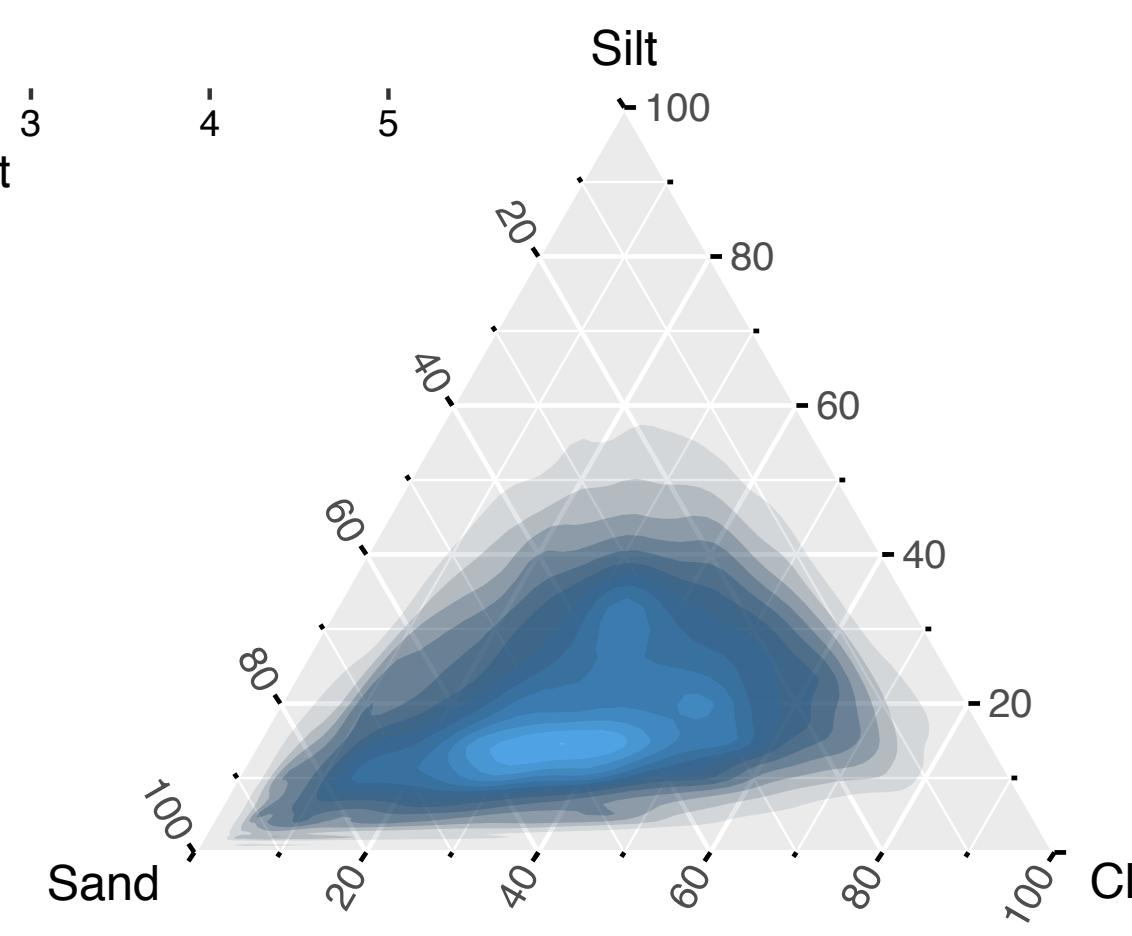
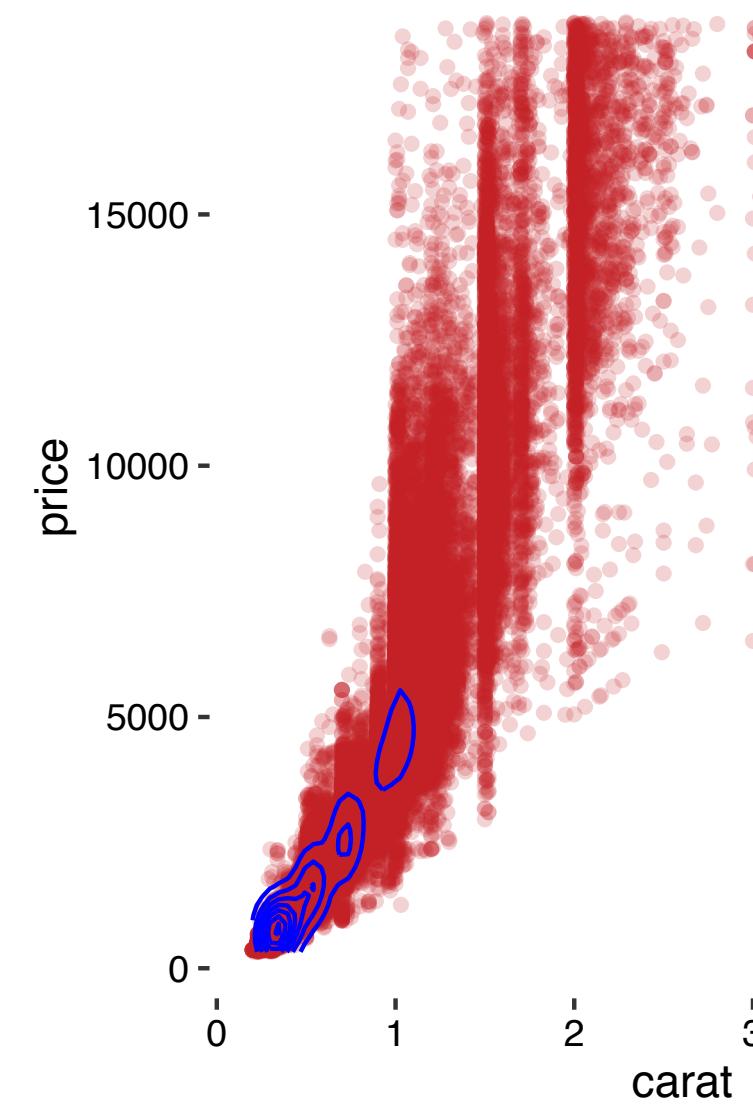
DATA VISUALIZATION WITH GGPLOT2

Introduction

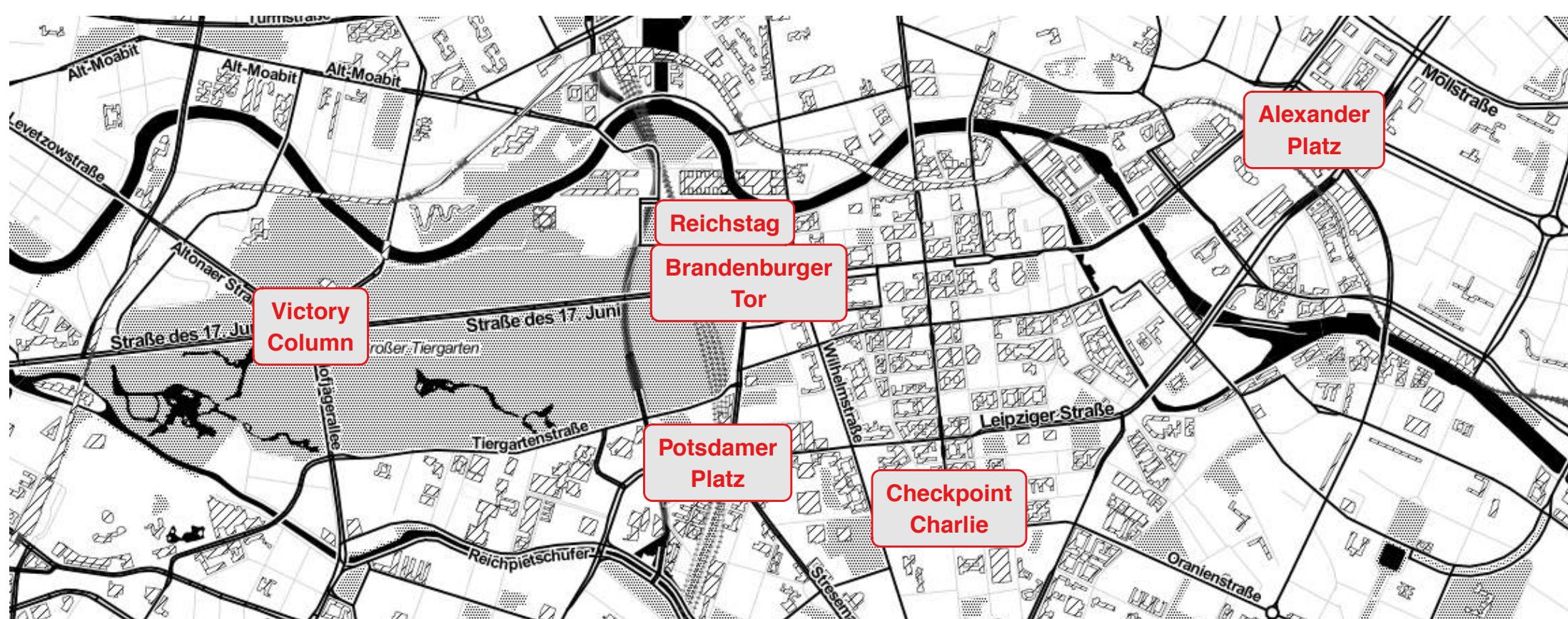
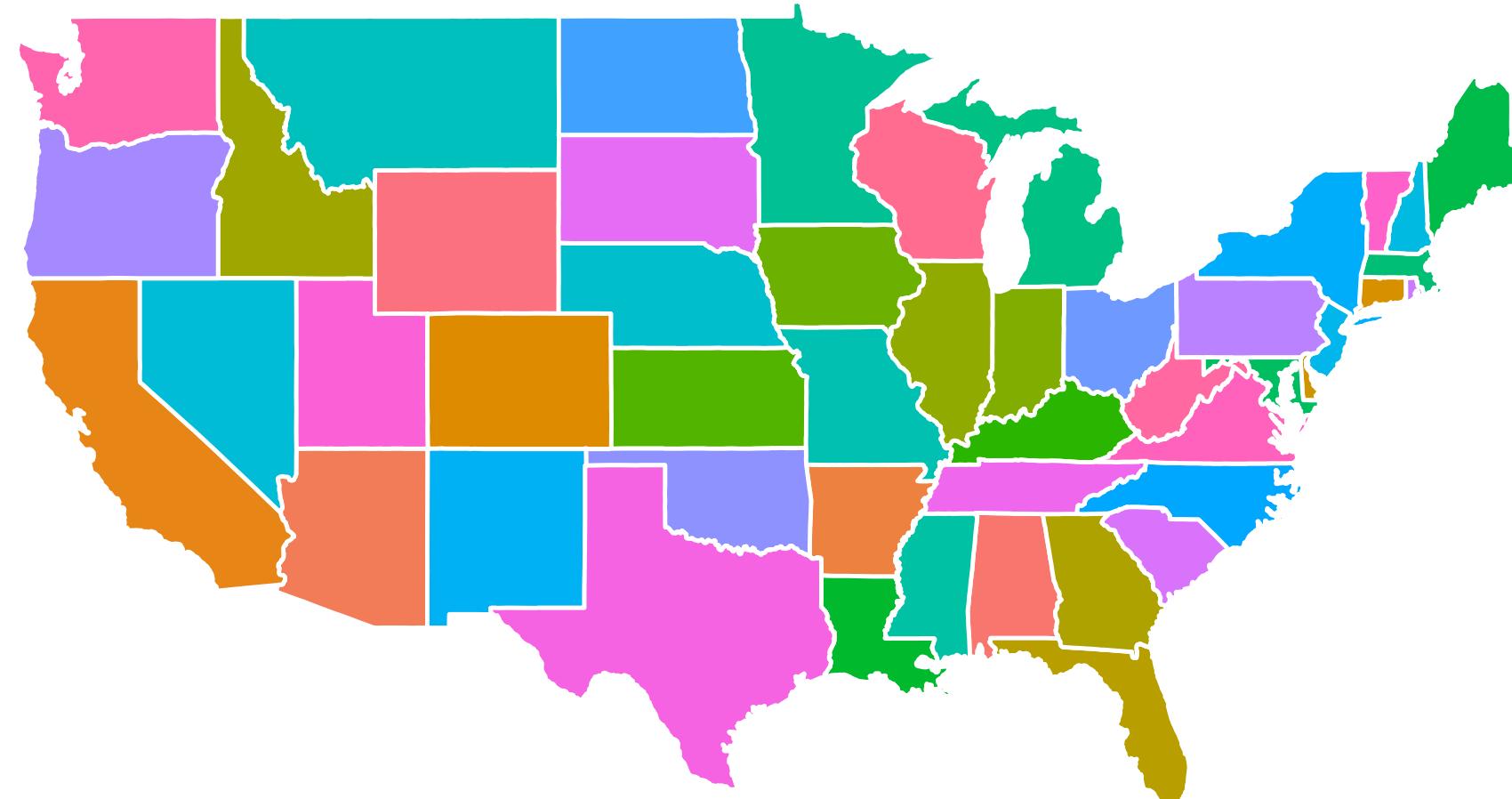
Chapter 1



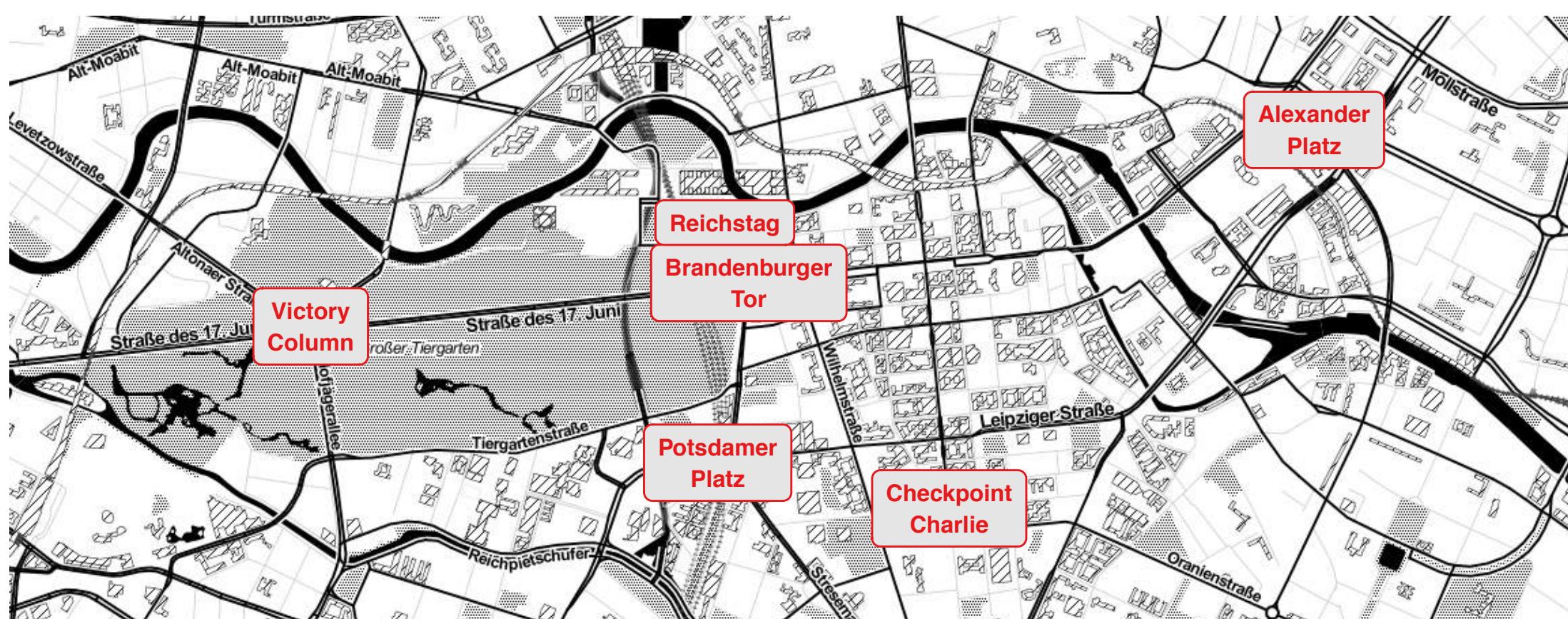
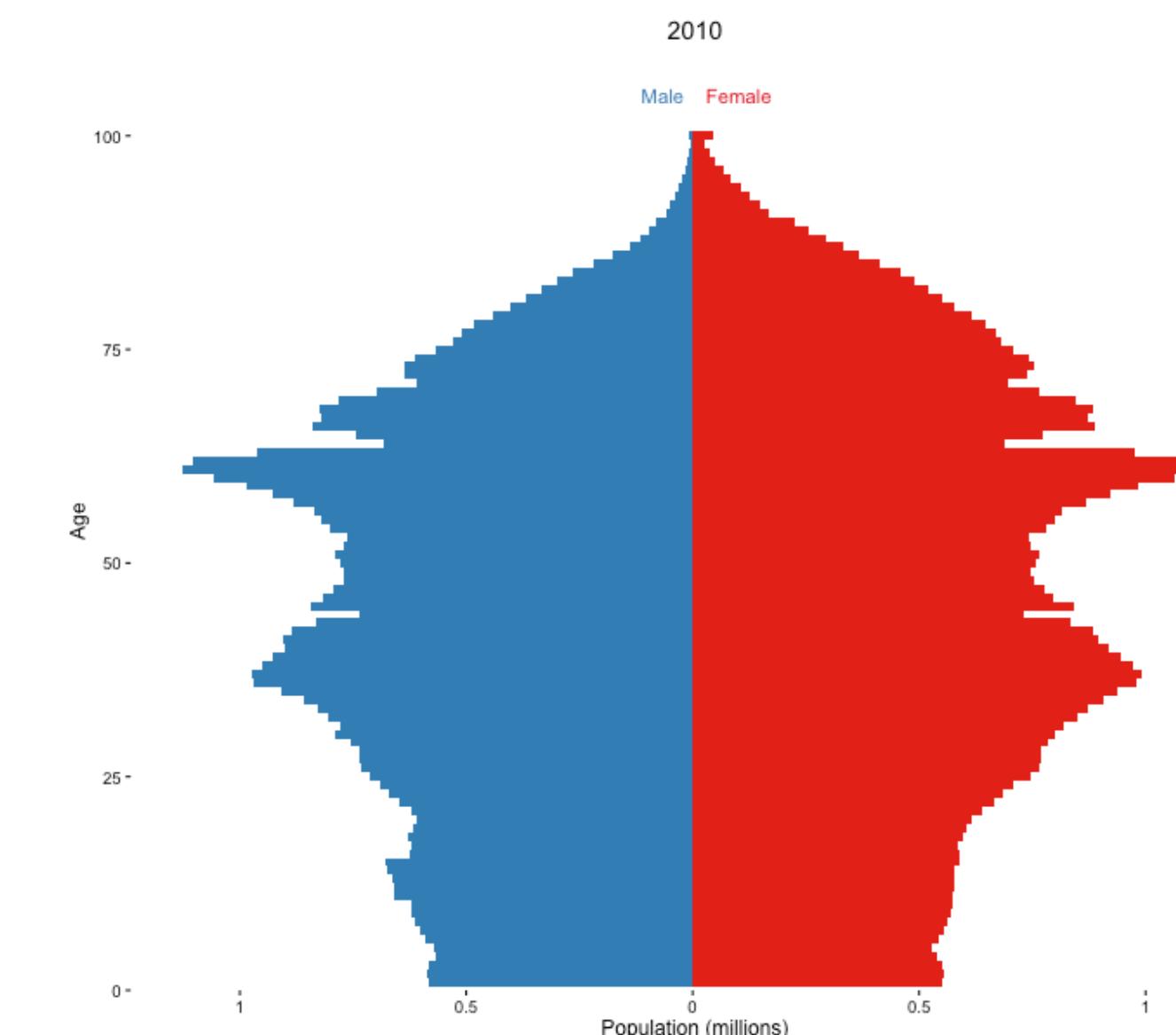
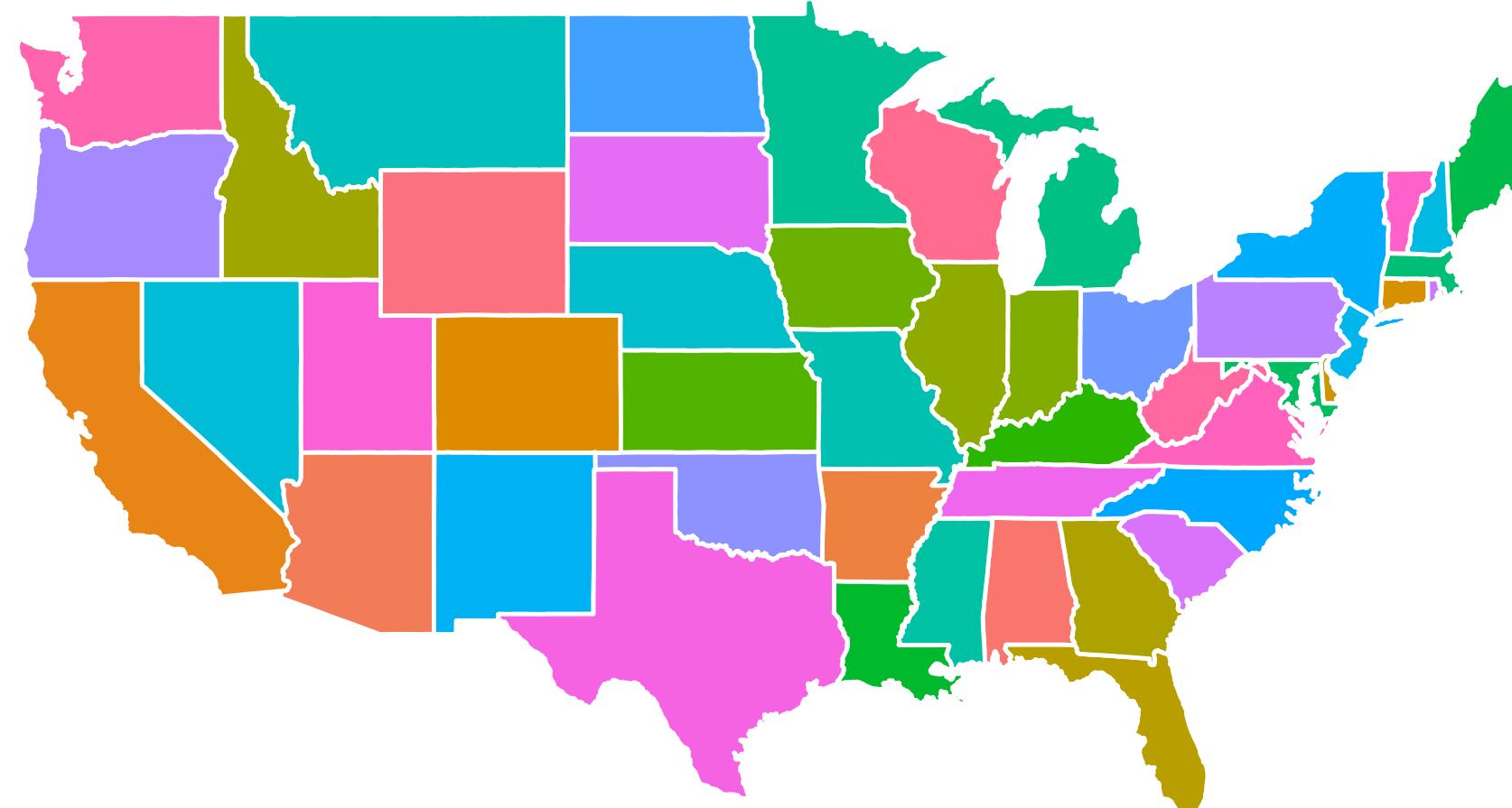
Chapter 2



Chapter 3



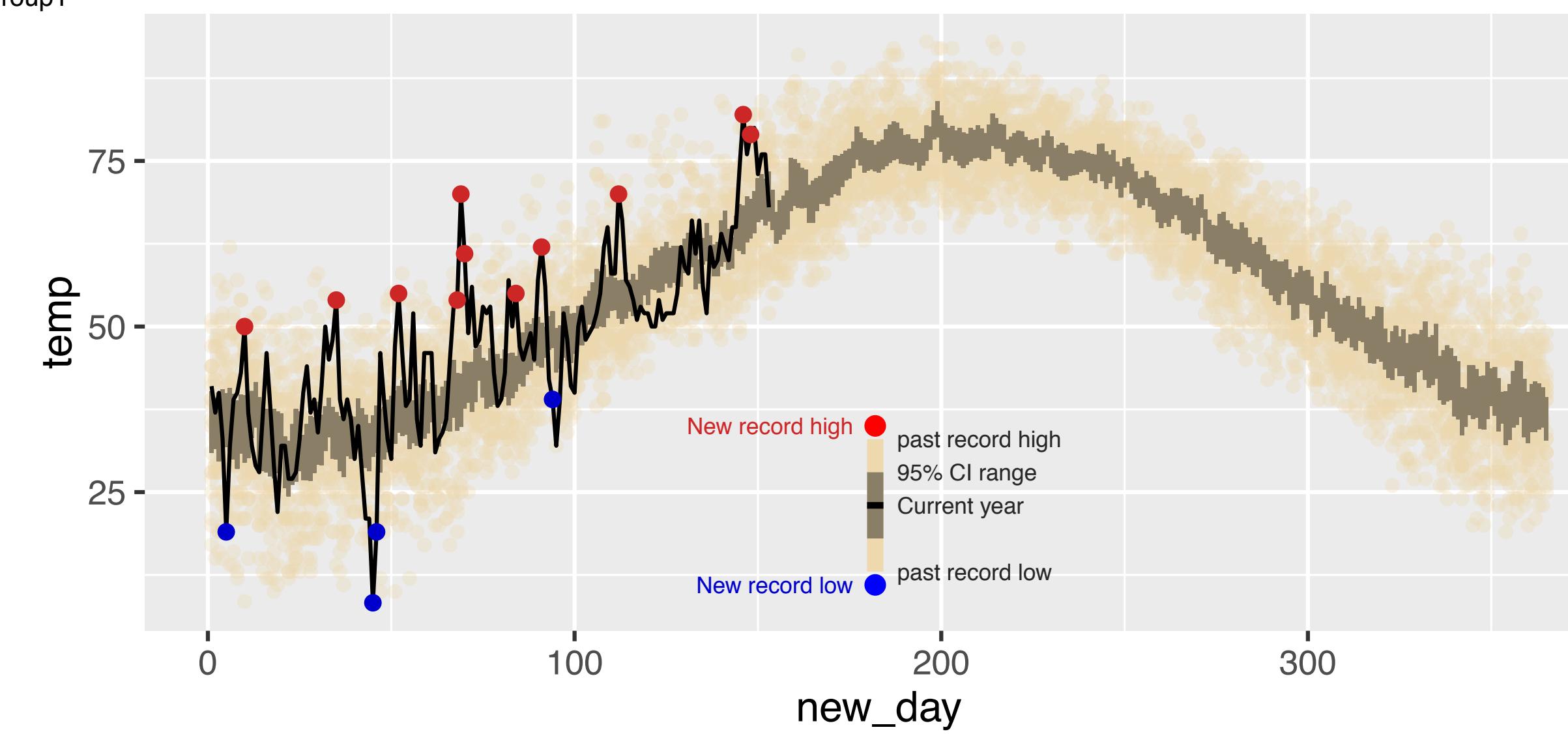
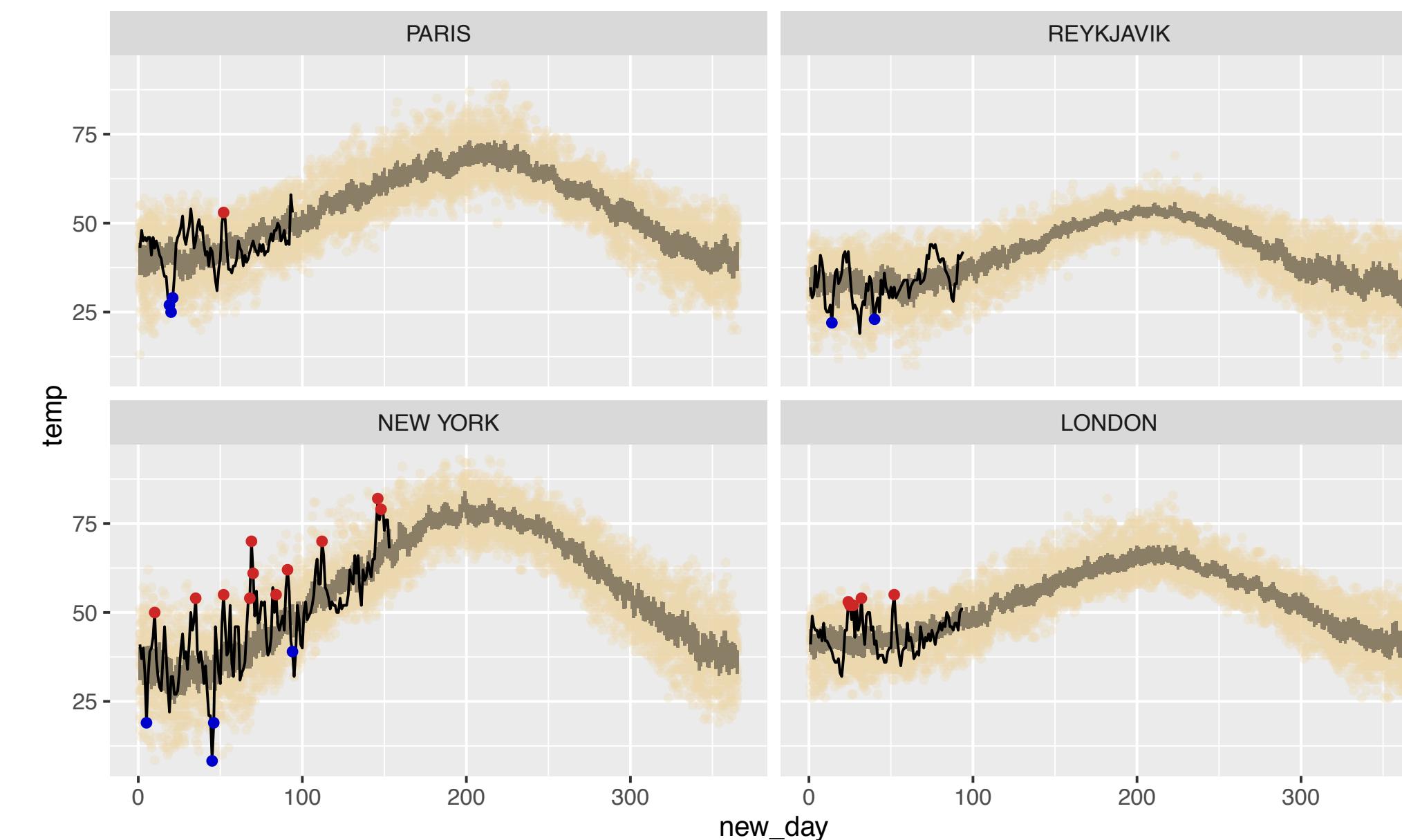
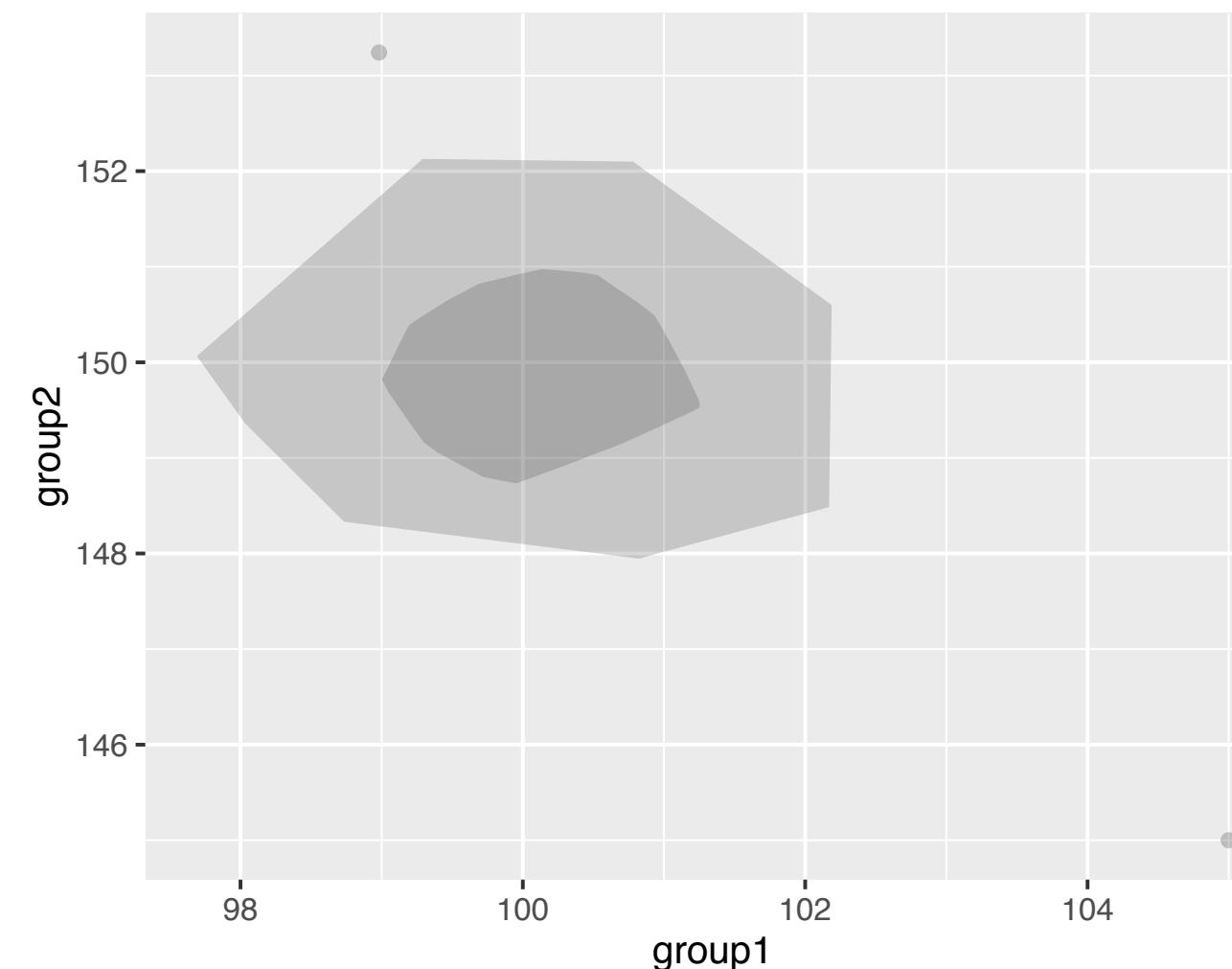
Chapter 3



Chapter 4

- Introduction to grid
- Manipulating graphical objects
- `ggplot_build()`
- `gridExtra`

Chapter 5





DATA VISUALIZATION WITH GGPLOT2

Let's practice!



DATA VISUALIZATION WITH GGPLOT2

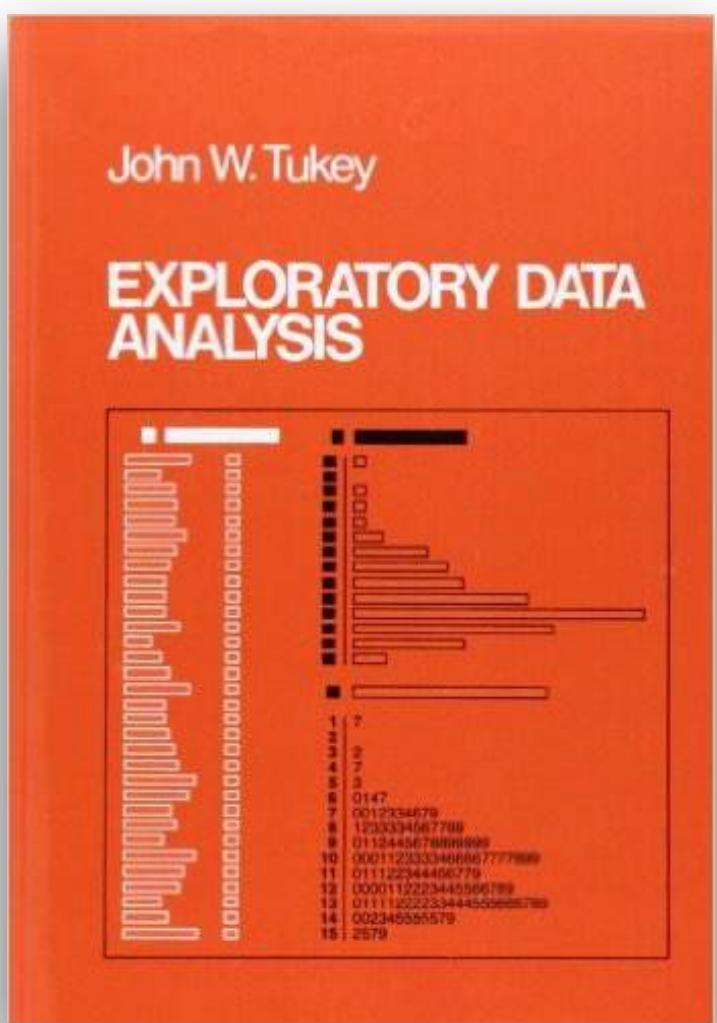
Box Plots

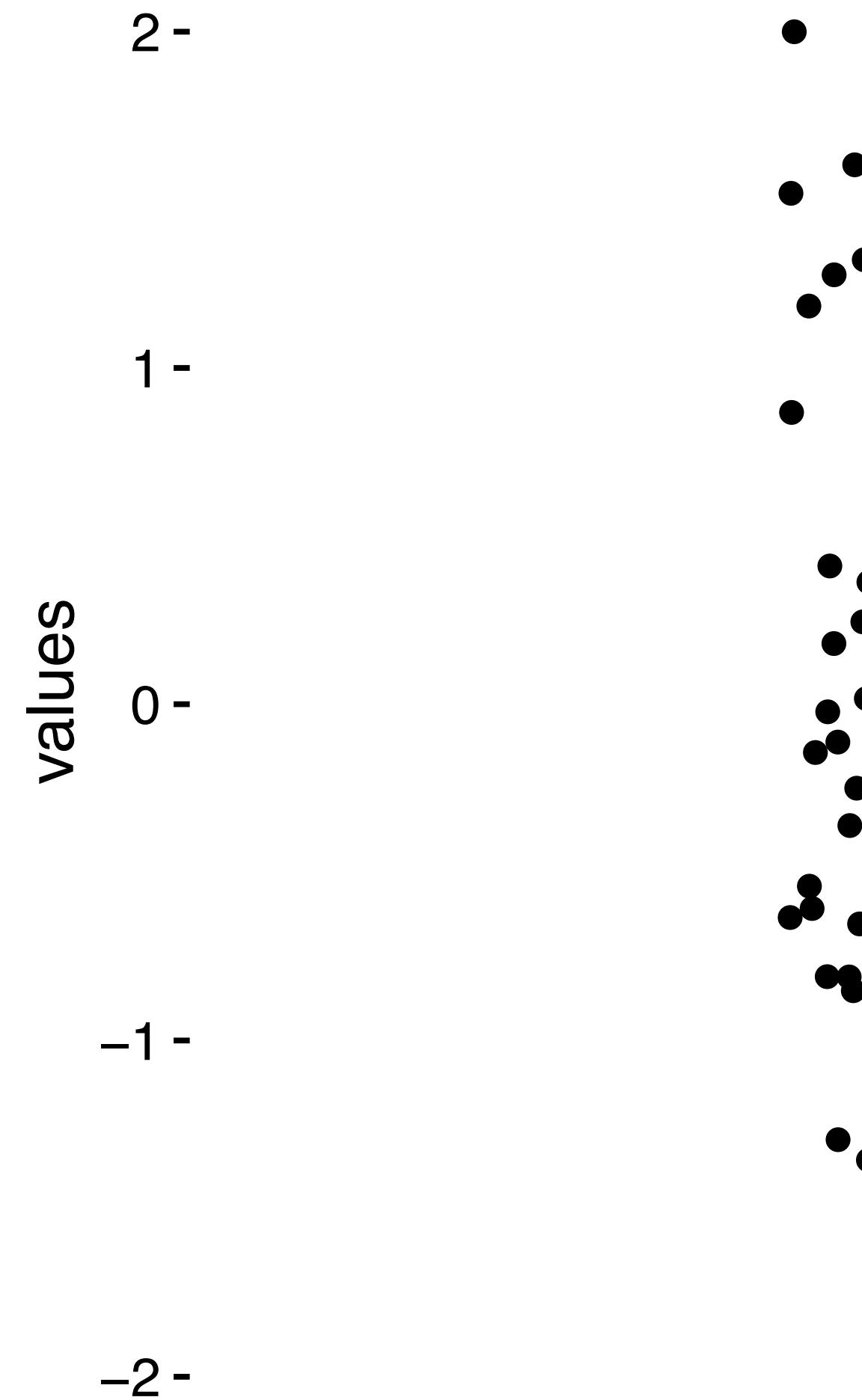
Statistical plots

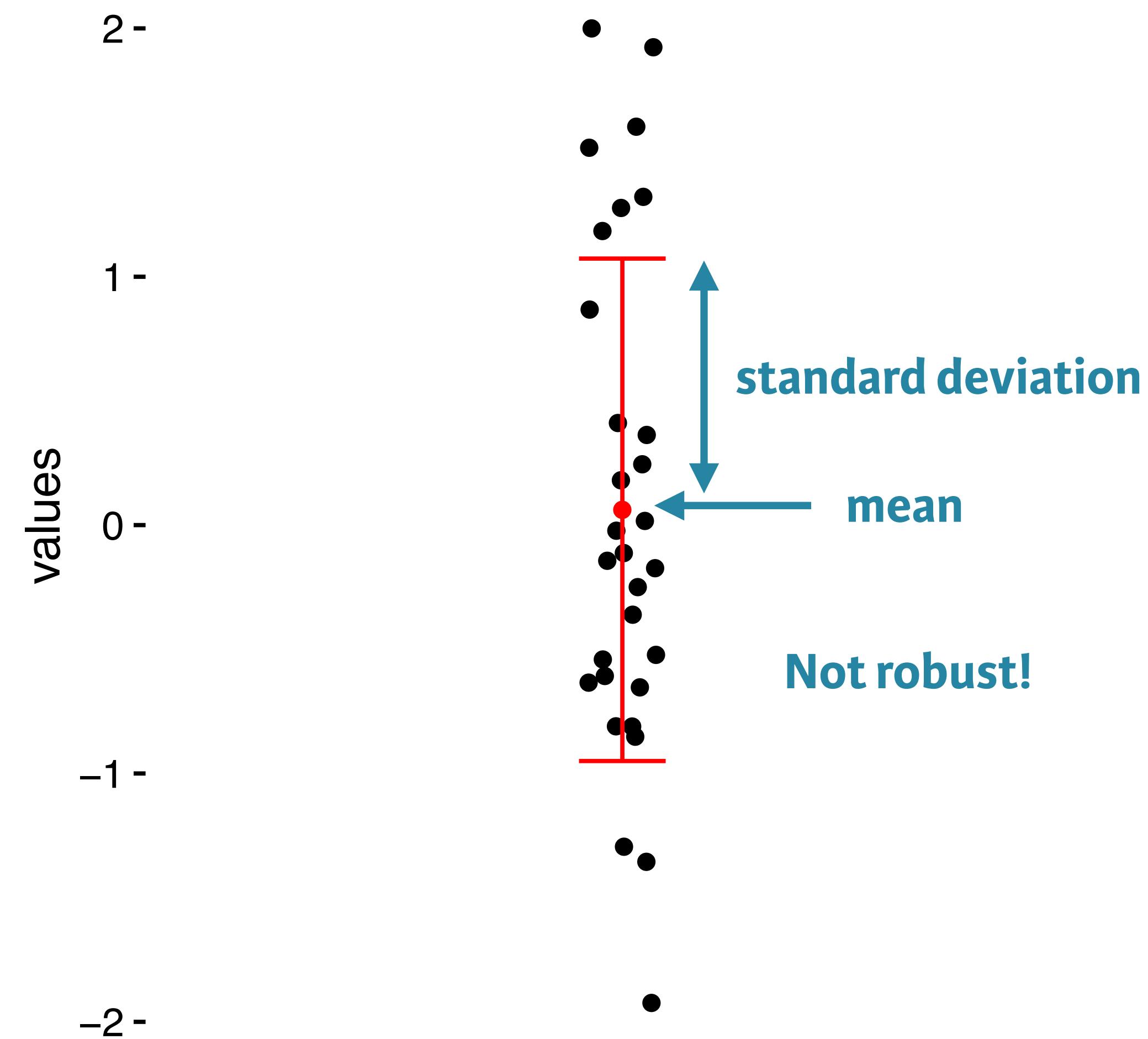
- Academic audience
- 2 common types
 - Box plots
 - Density plots
- Case study: 2D box plots

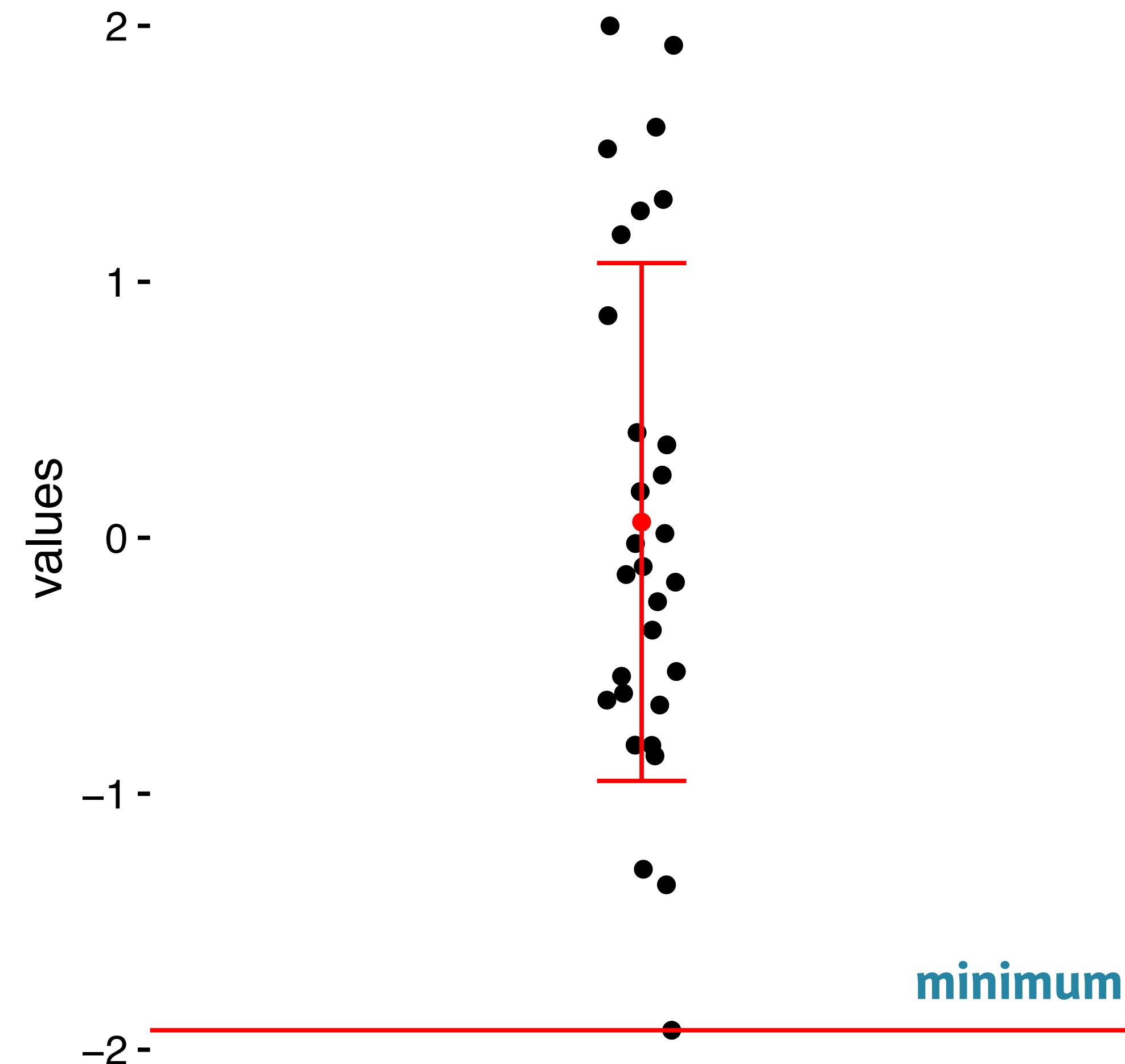
Box plot

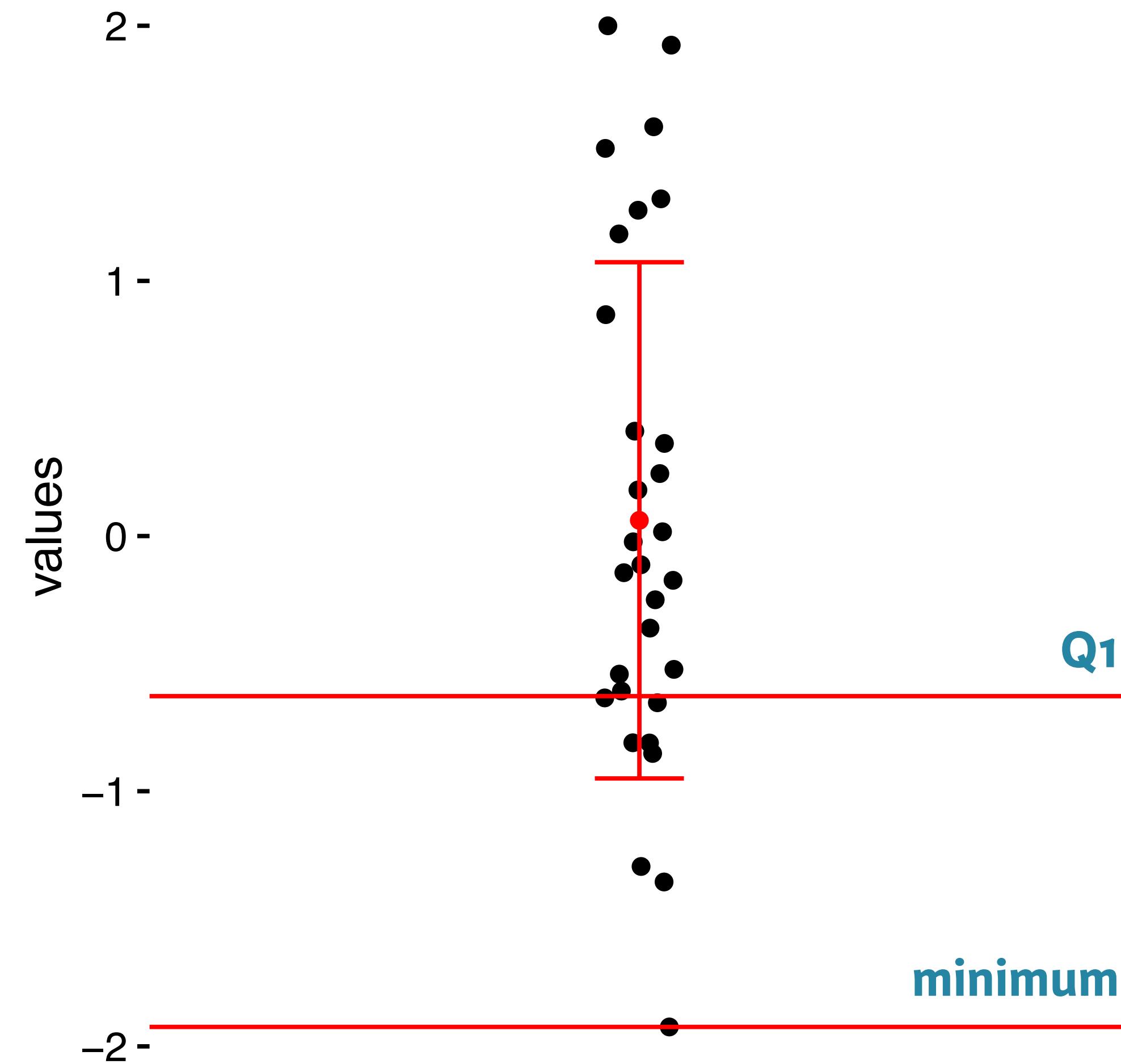
- John Tukey - Exploratory Data Analysis
- Visualizing the 5 number summary

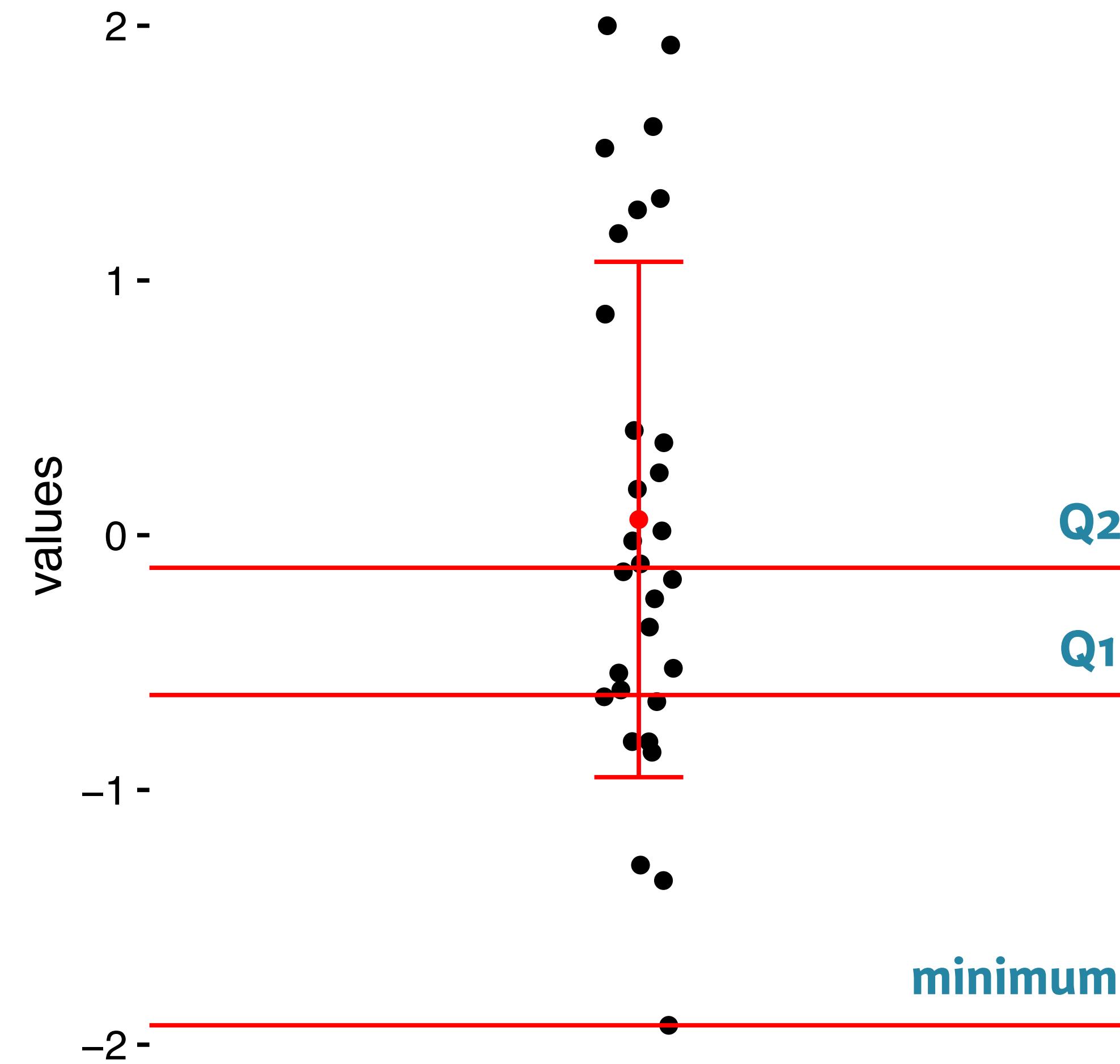


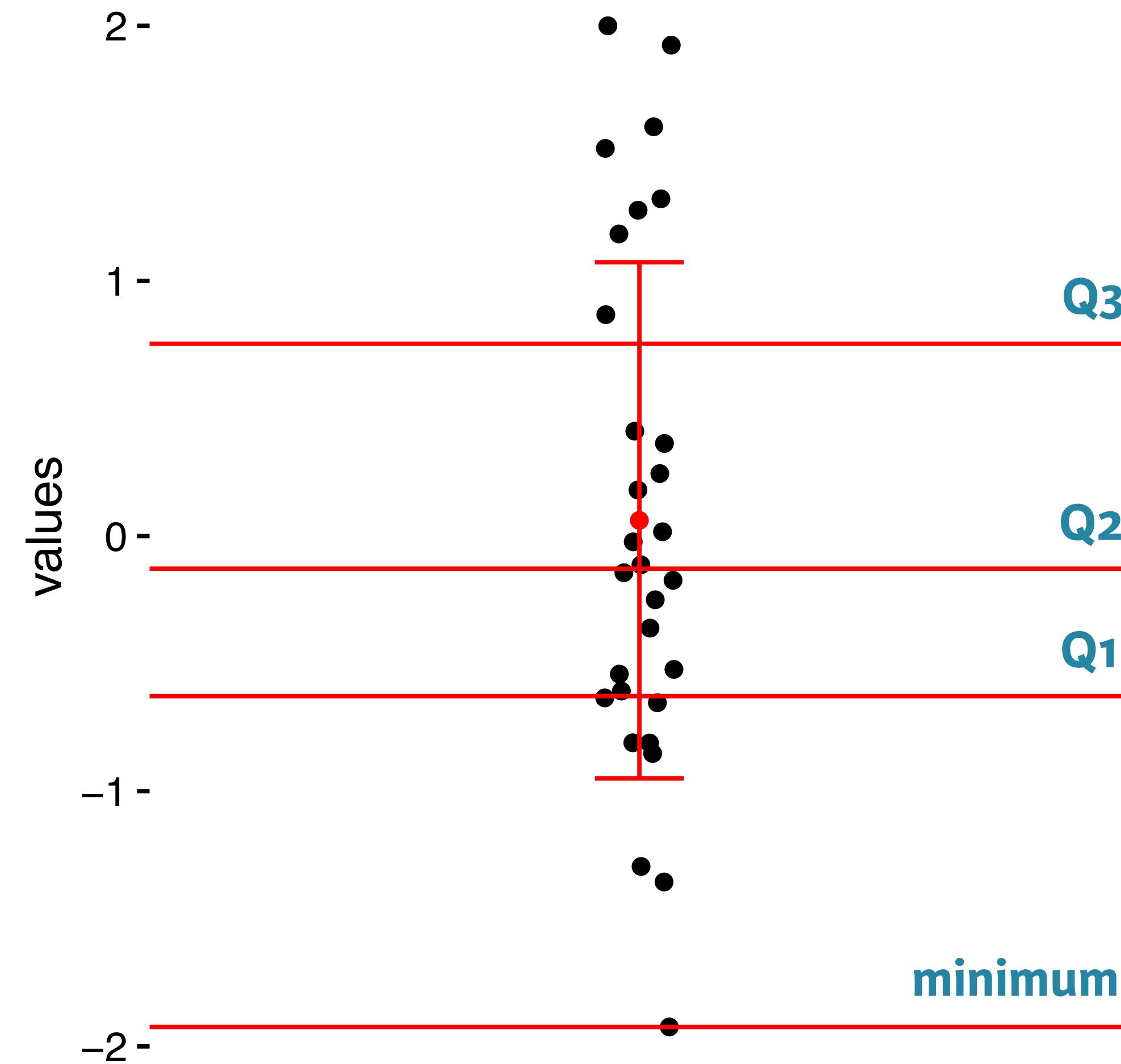


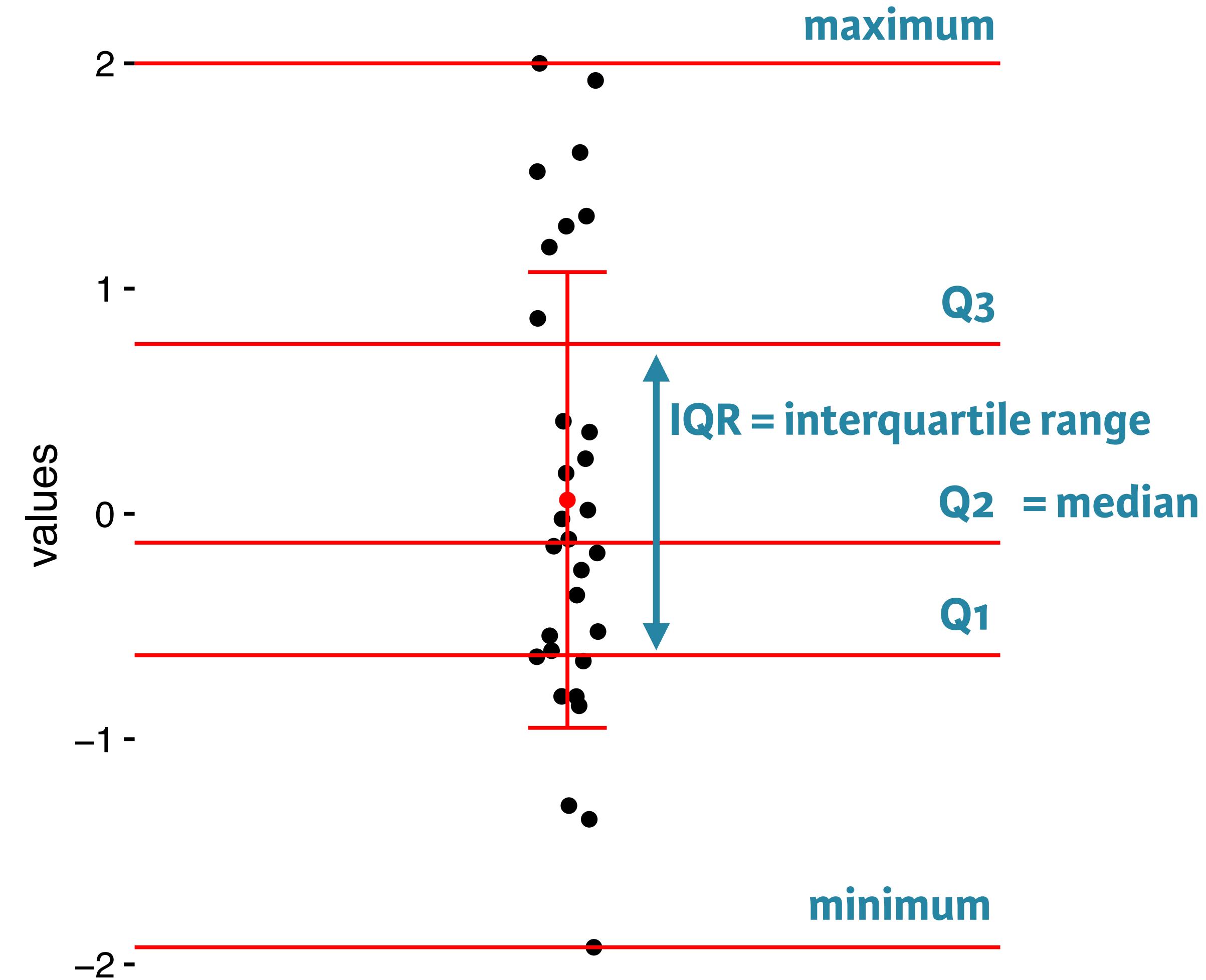


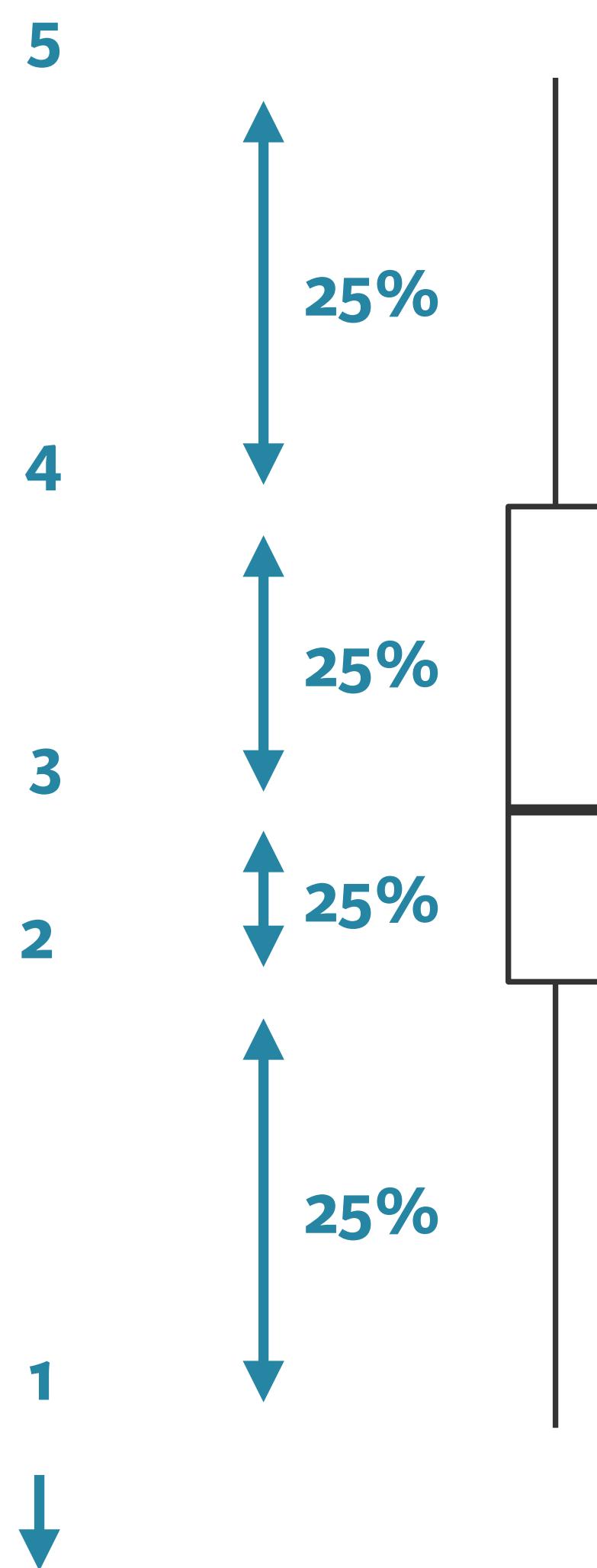
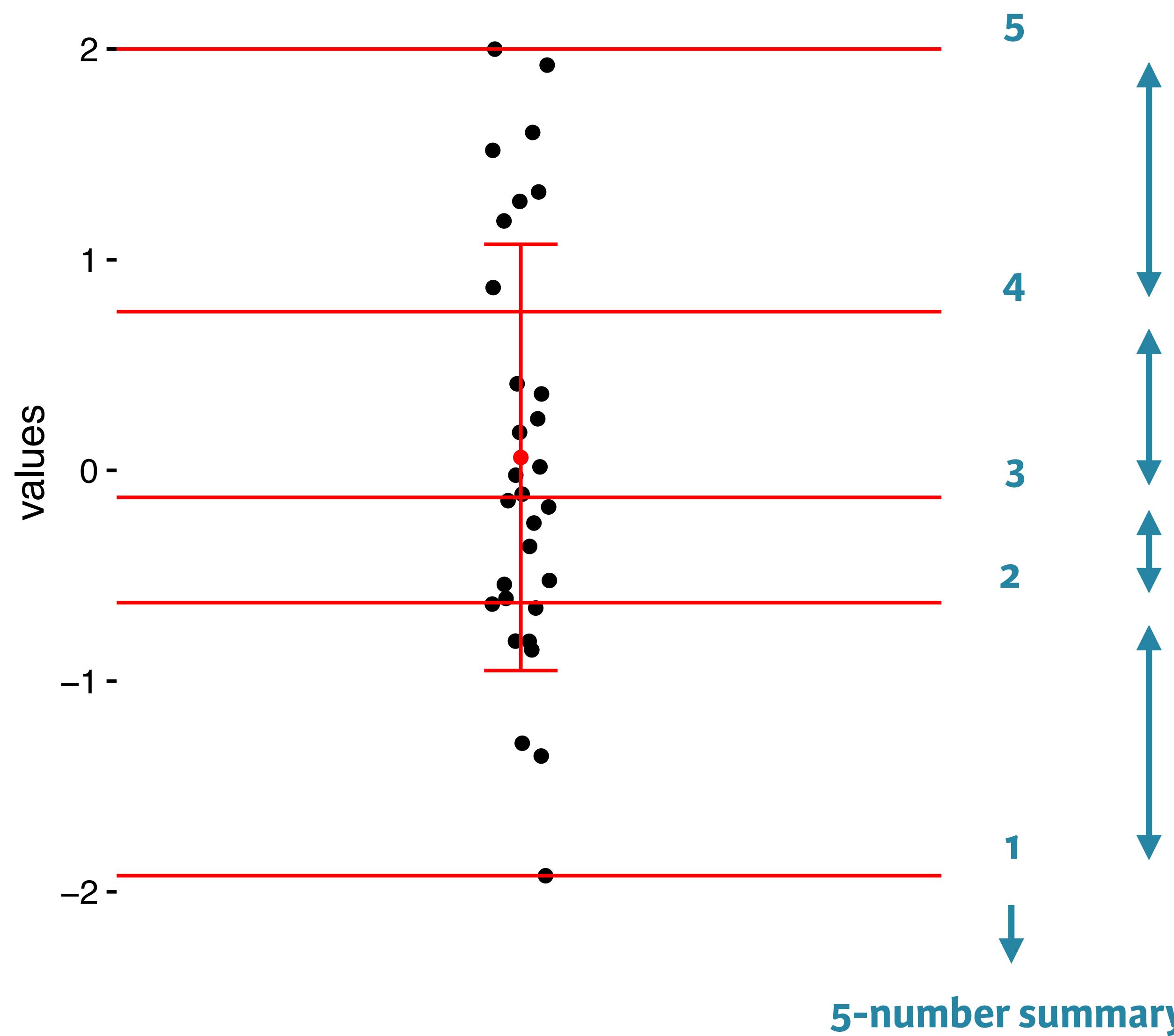




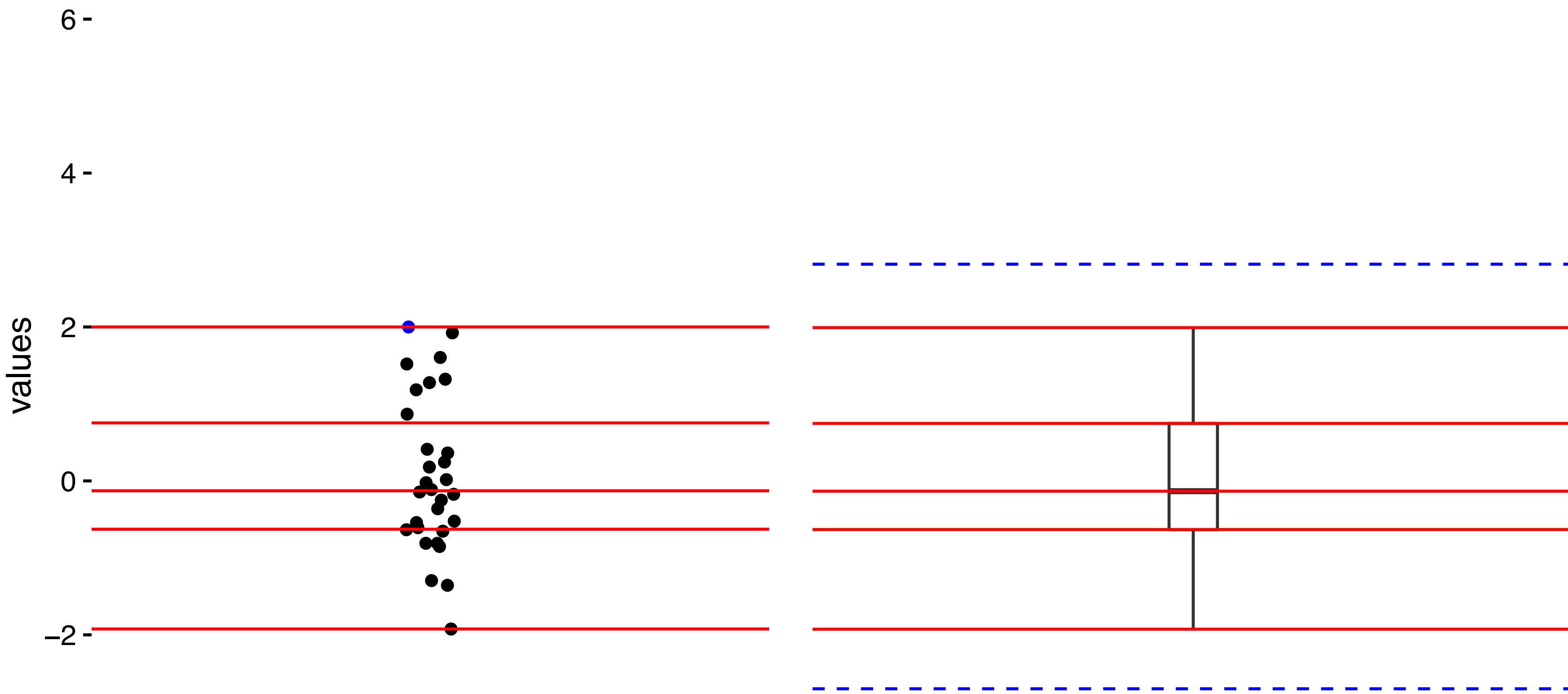


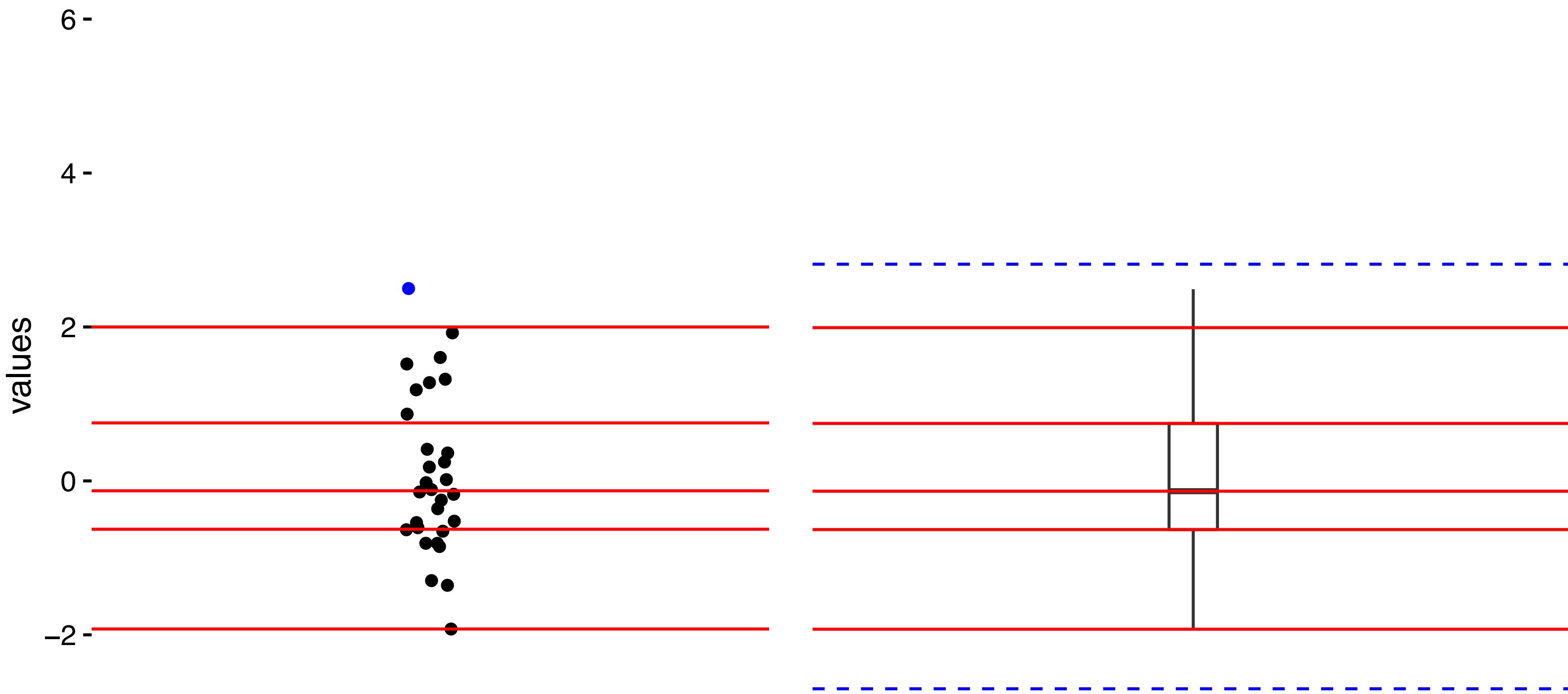


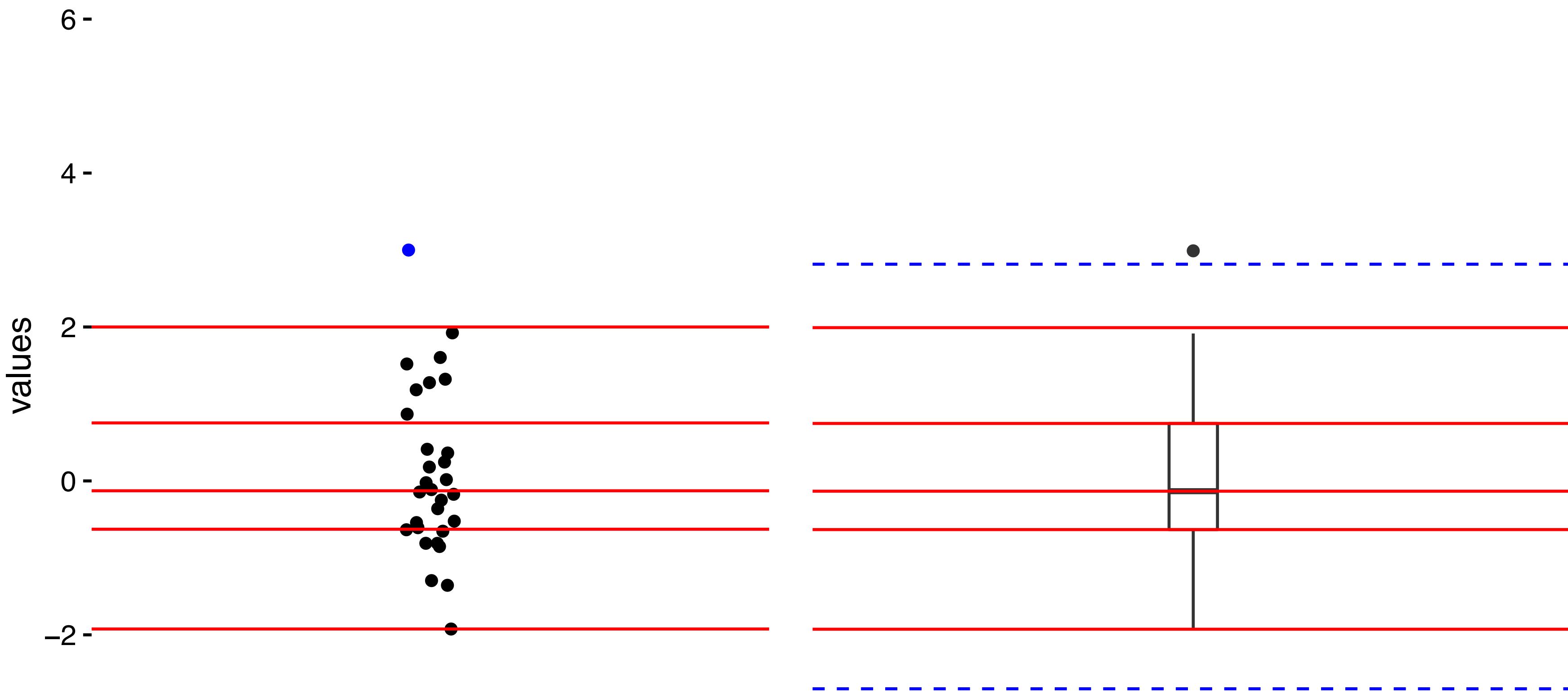


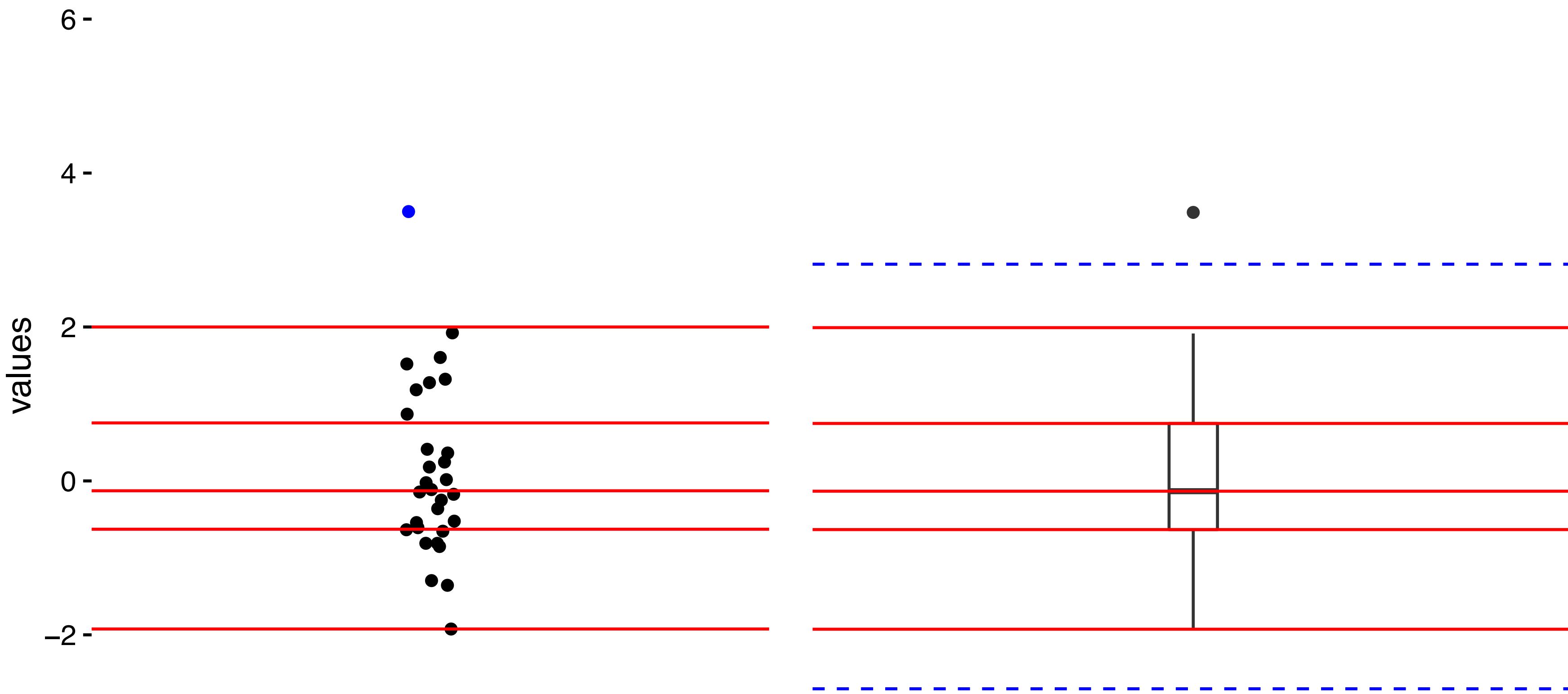


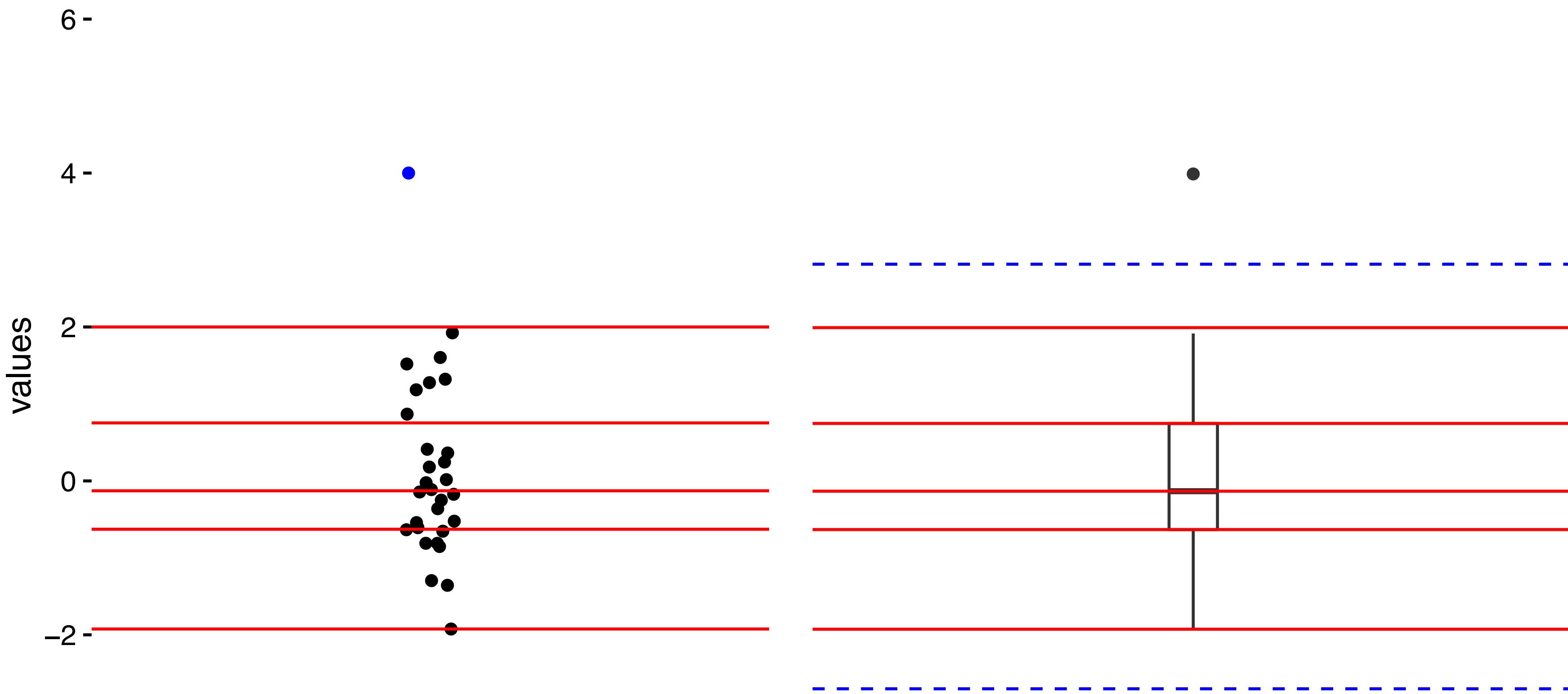
5-number summary

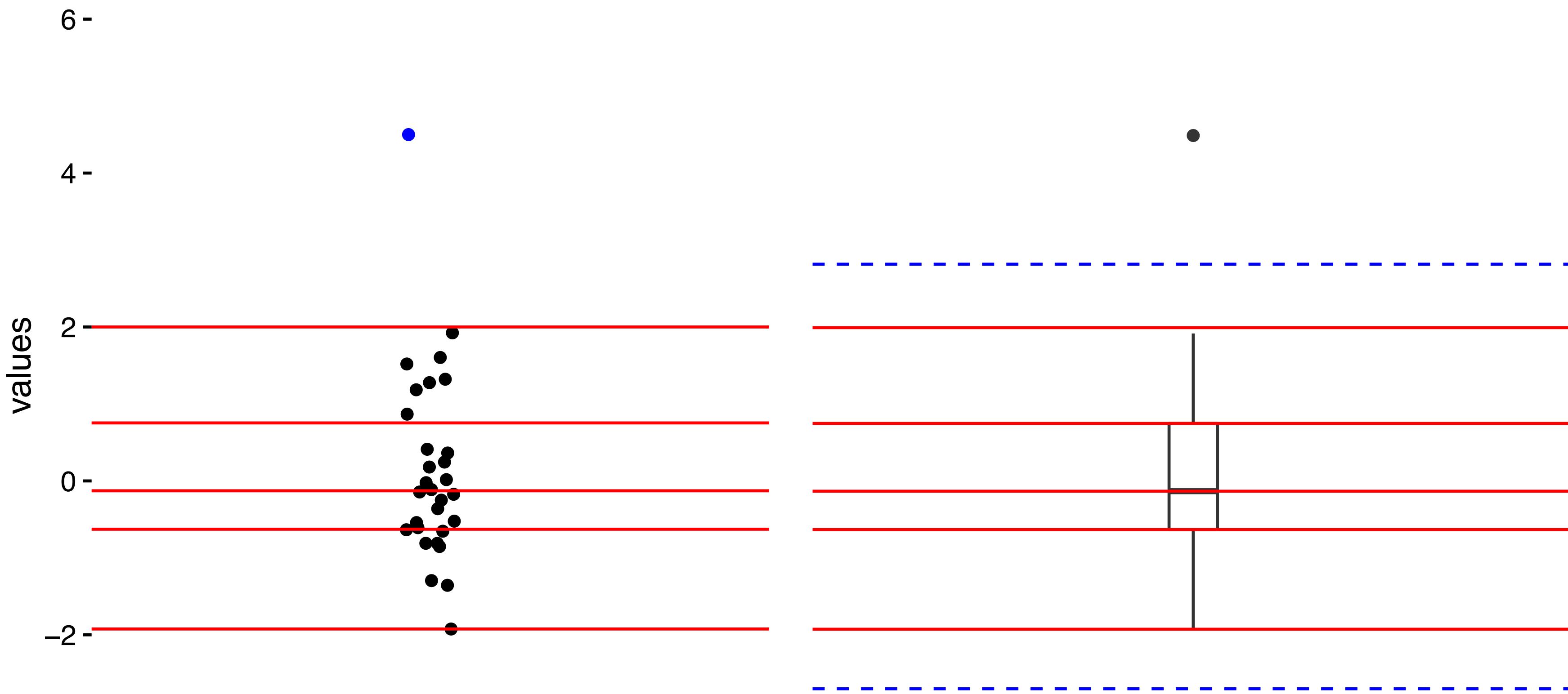


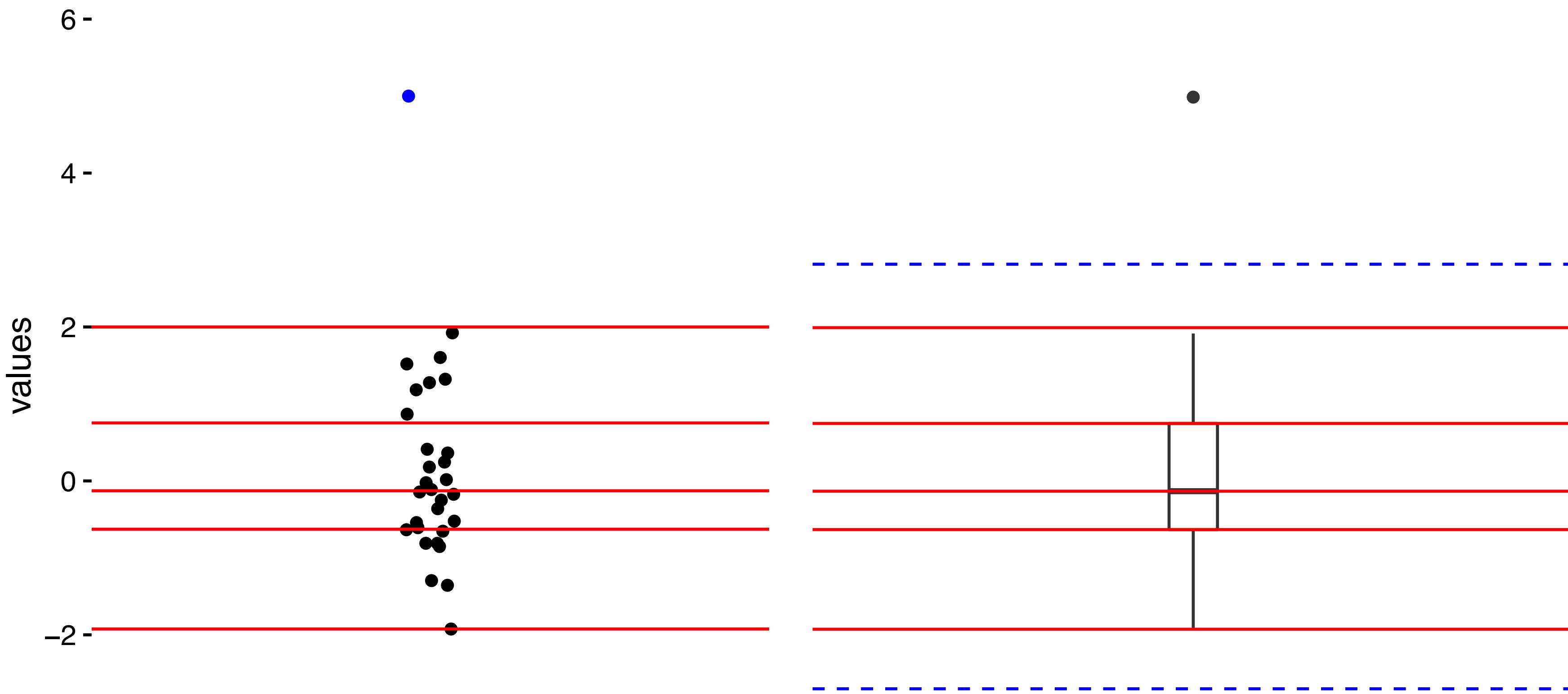














DATA VISUALIZATION WITH GGPLOT2

Let's practice!

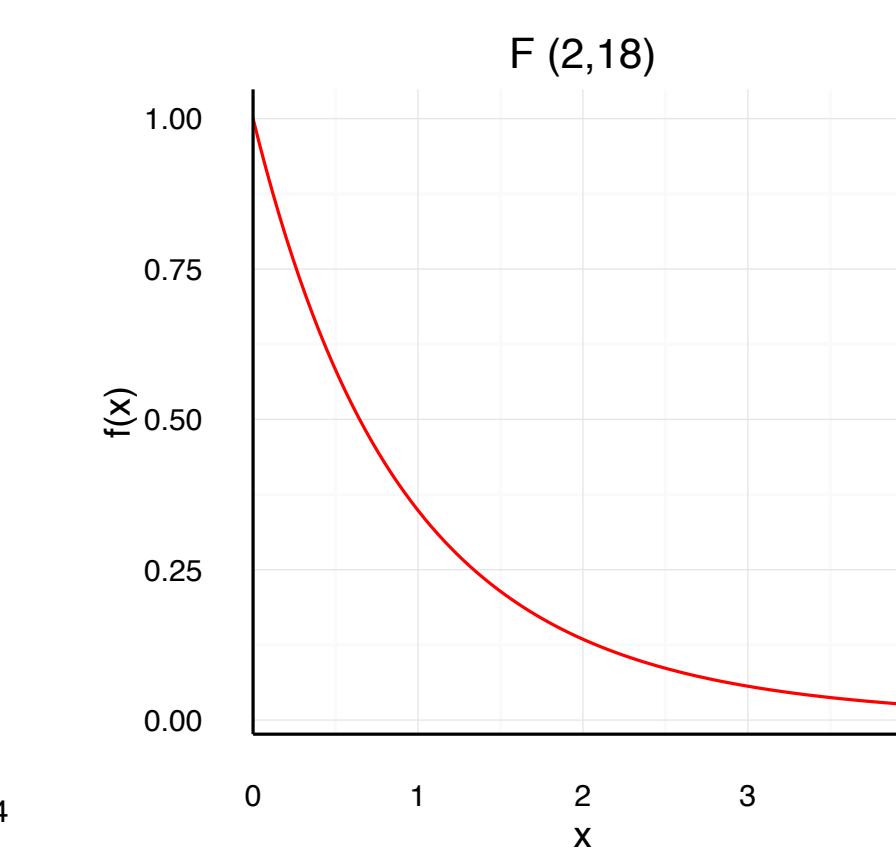
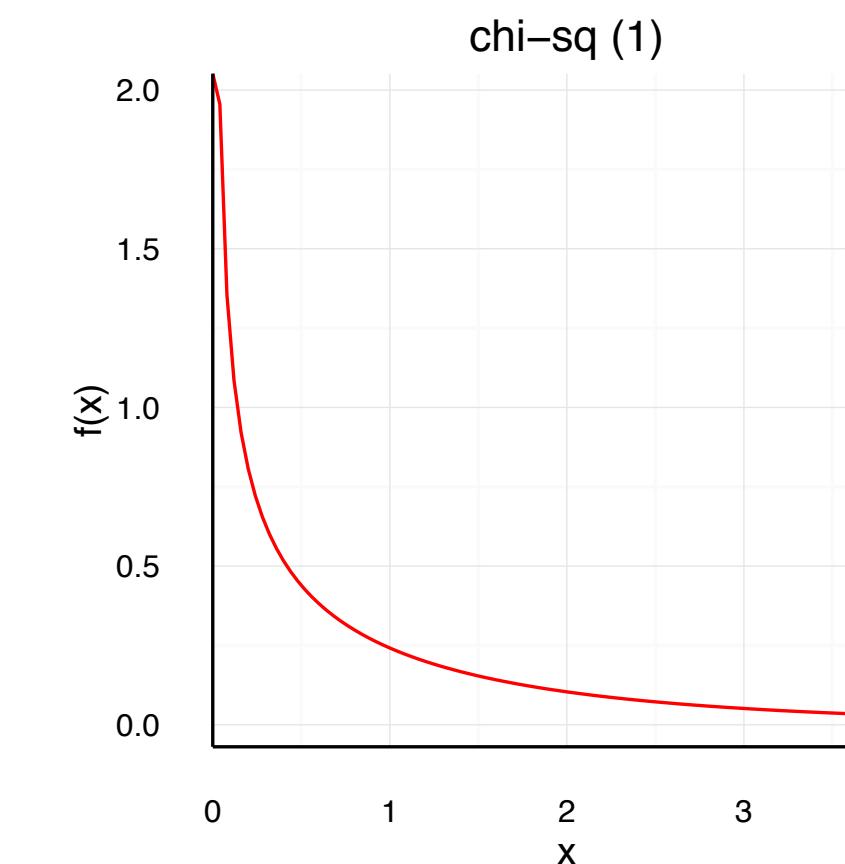
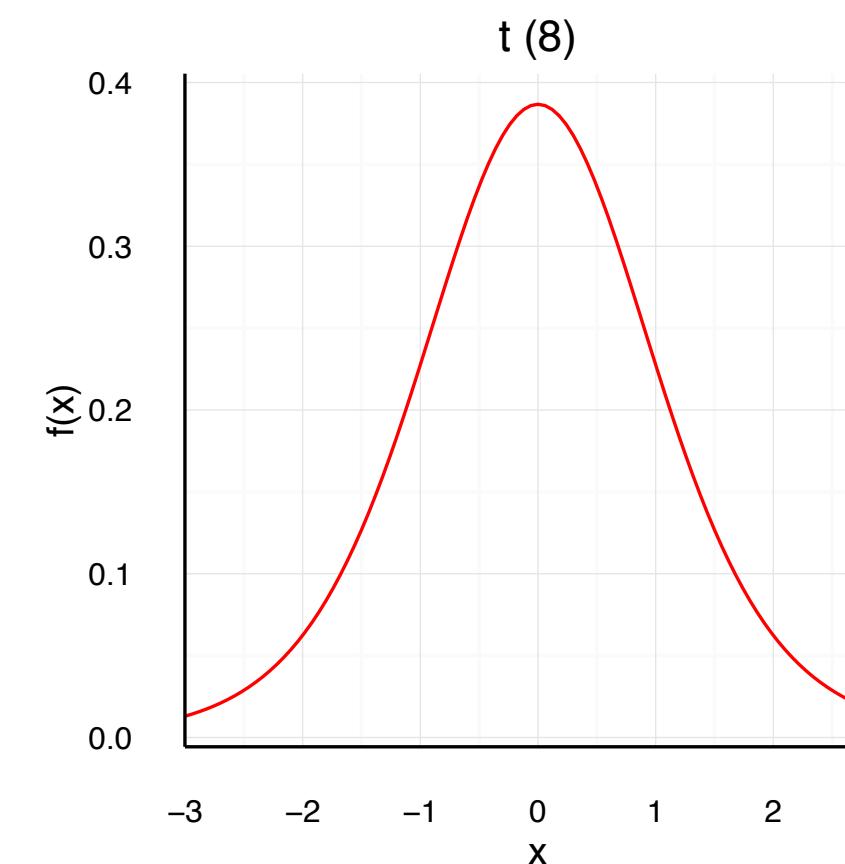
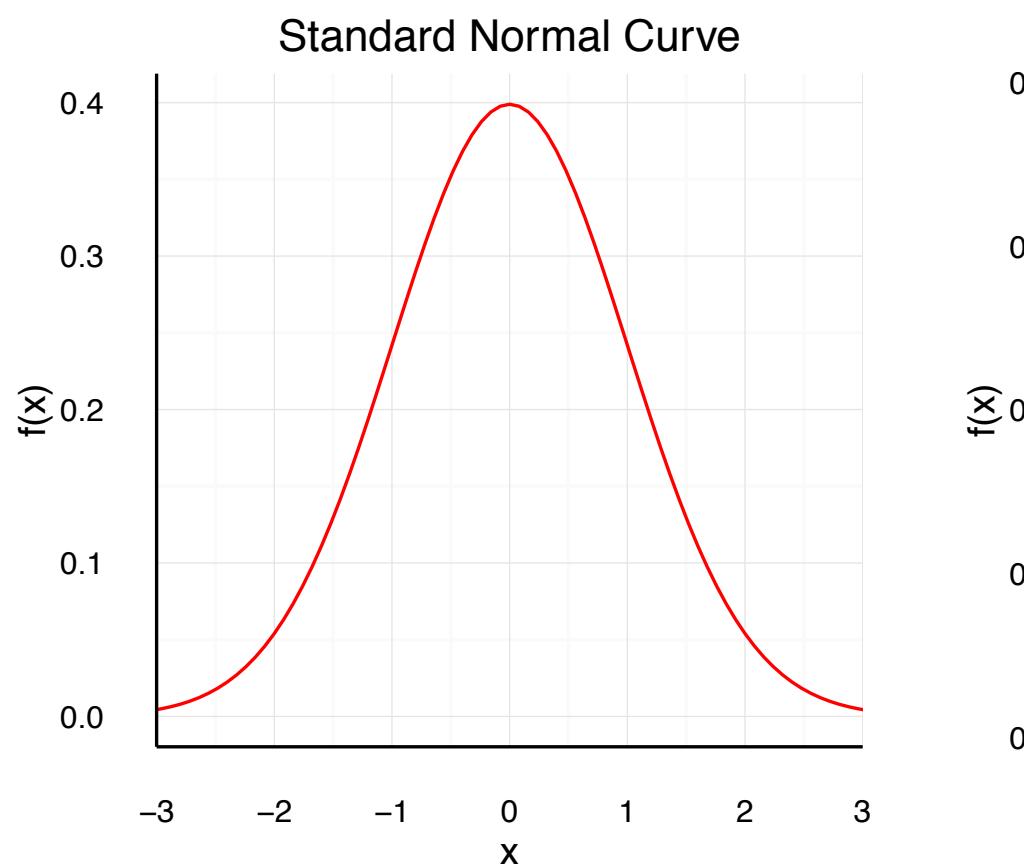


DATA VISUALIZATION WITH GGPLOT2

Density Plots

Density plot

- Distribution of univariate data
- Statistics
 - Probability Density Function



- Theoretical: based on formula
- Empirical: based on data

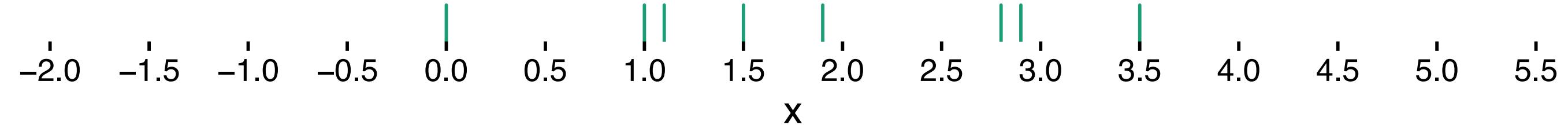
Kernel Density Estimate (KDE)

A sum of 'bumps' placed at the observations.

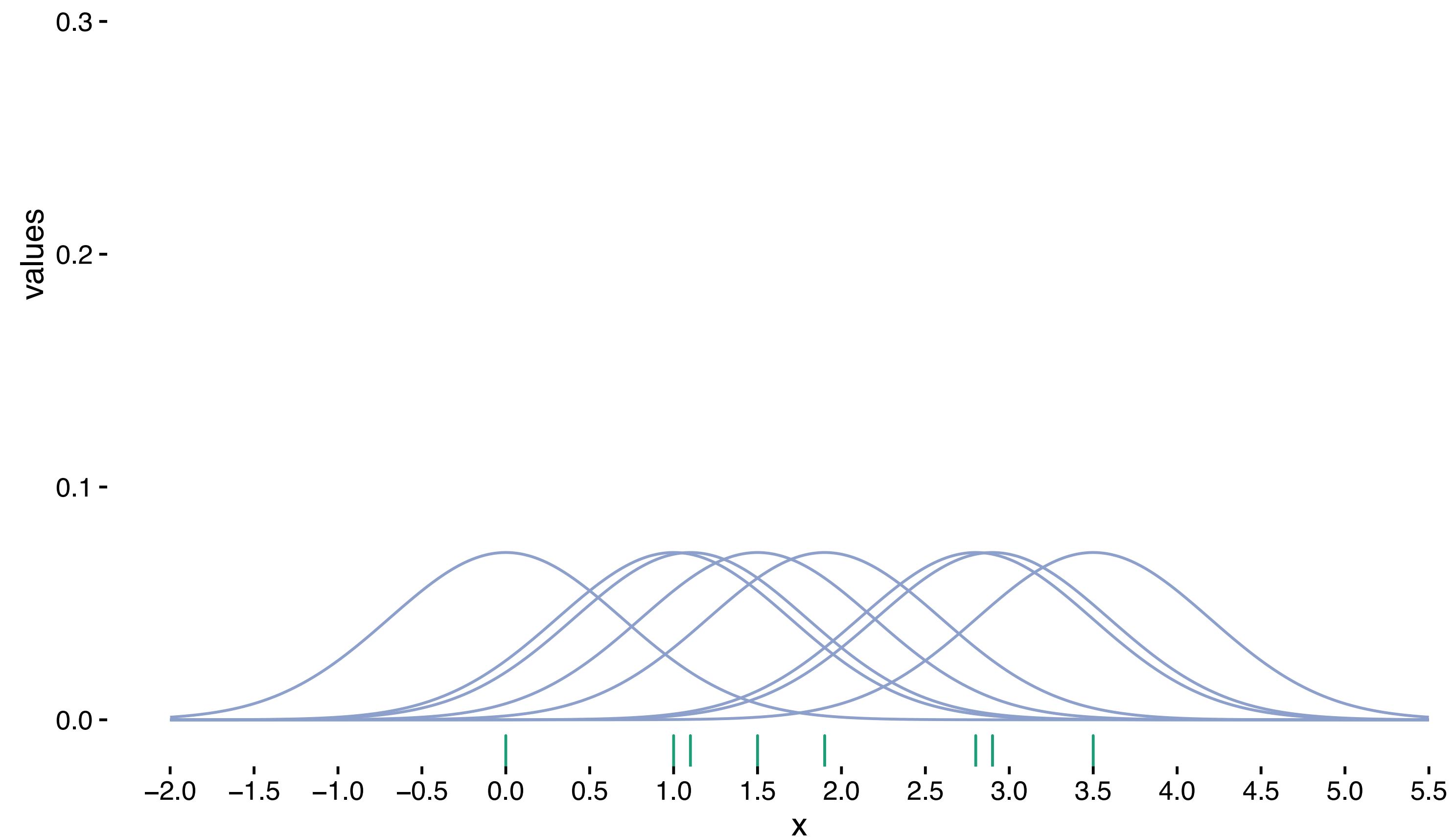
*The kernel function determines the shape of the bumps
while the window width, h , determines their width.*

Example

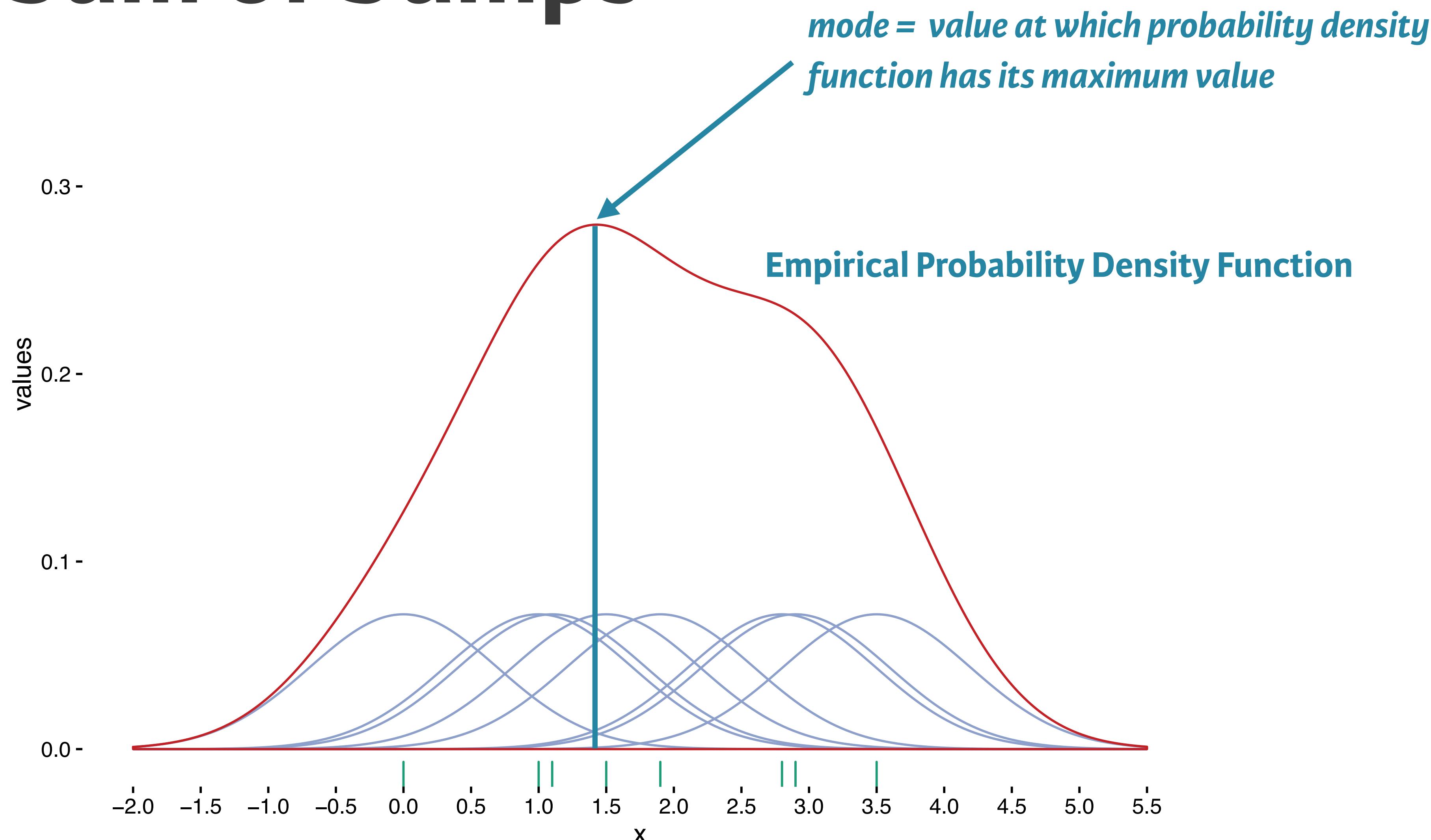
```
> x <- c(0.0, 1.0, 1.1, 1.5, 1.9, 2.8, 2.9, 3.5)  
  
> x  
[1] 0.0 1.0 1.1 1.5 1.9 2.8 2.9 3.5
```



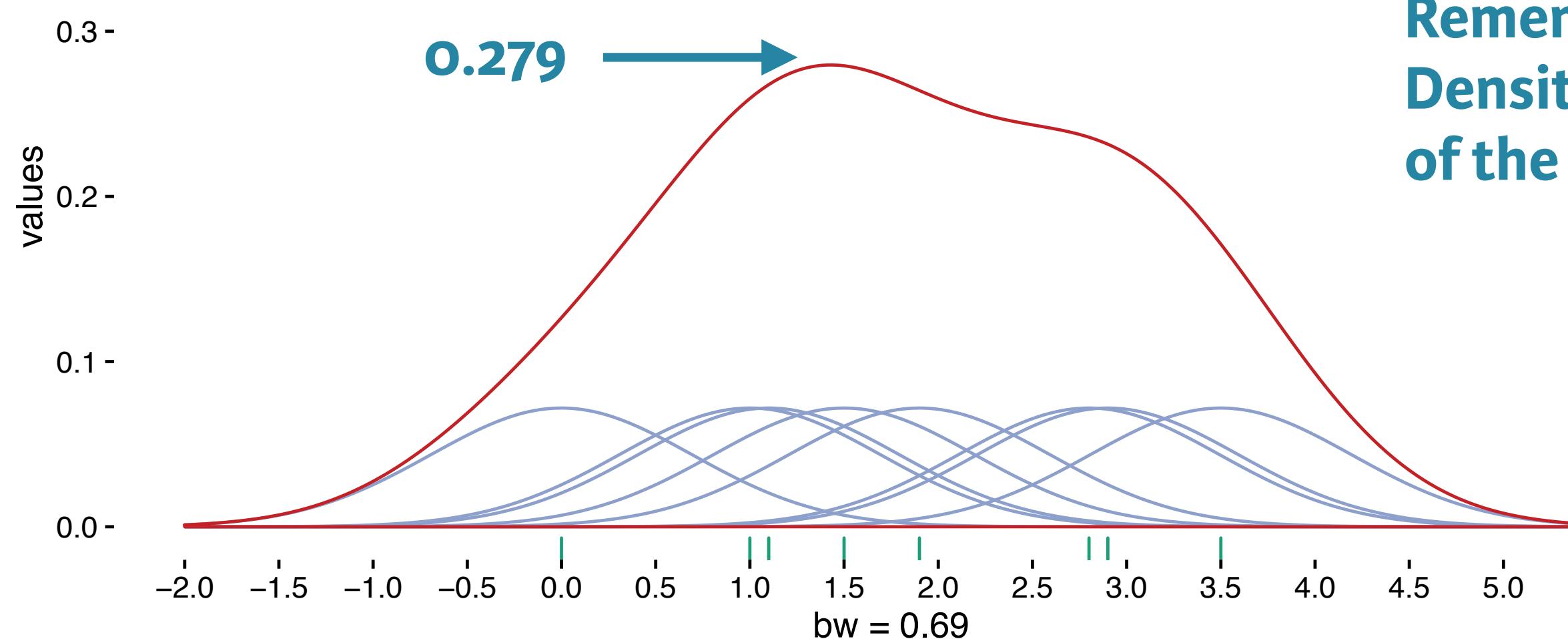
Bumps



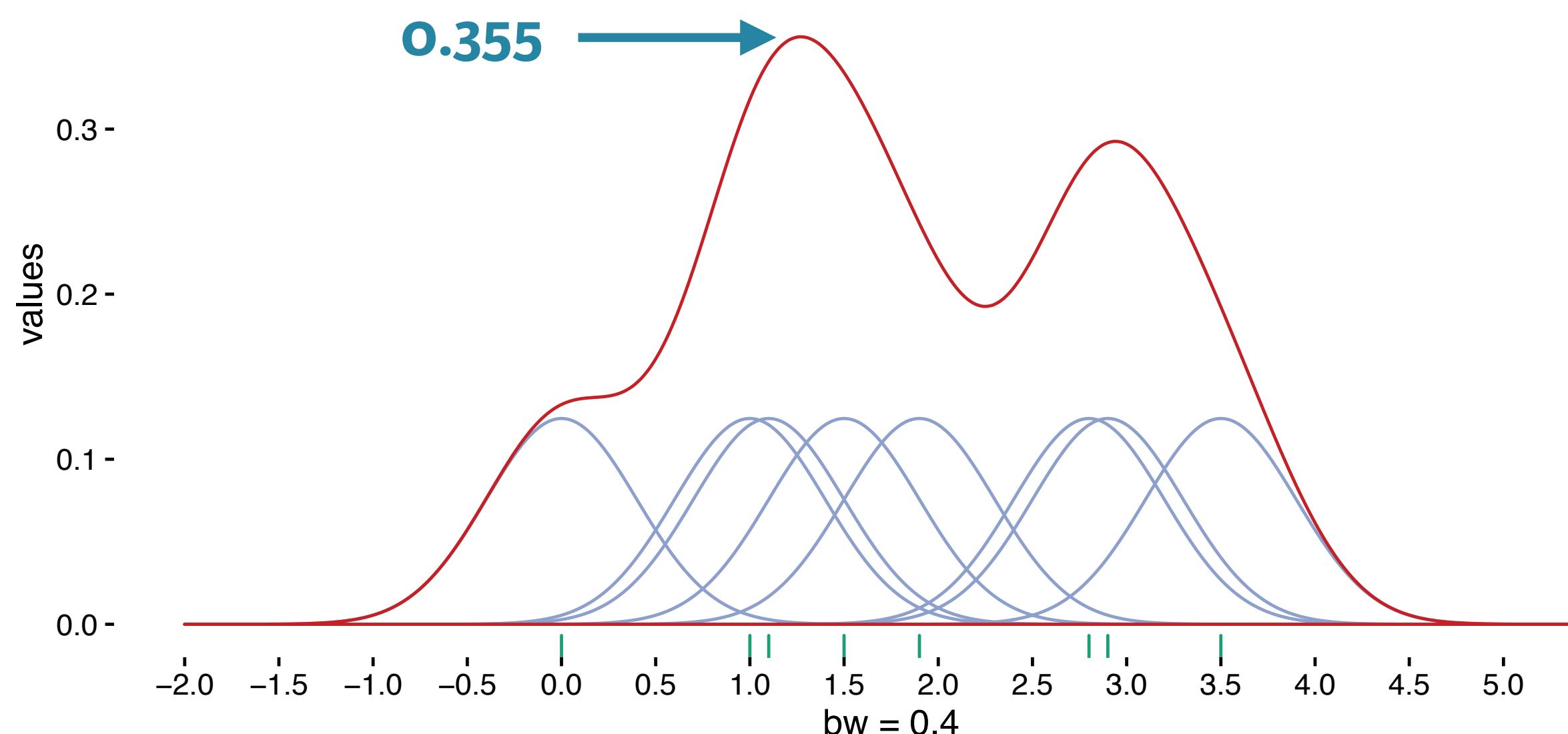
Sum of bumps

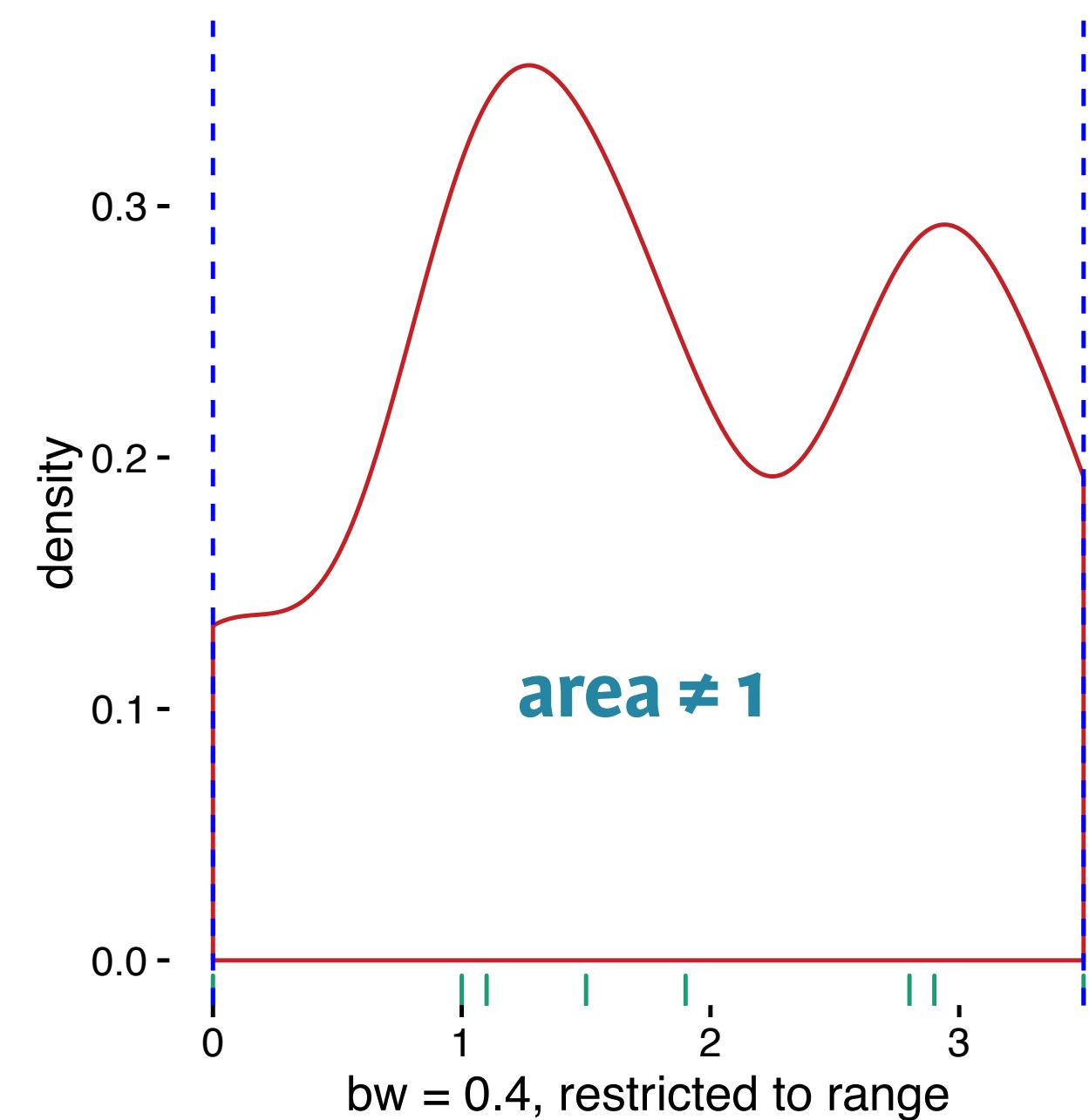


Bandwidth - h

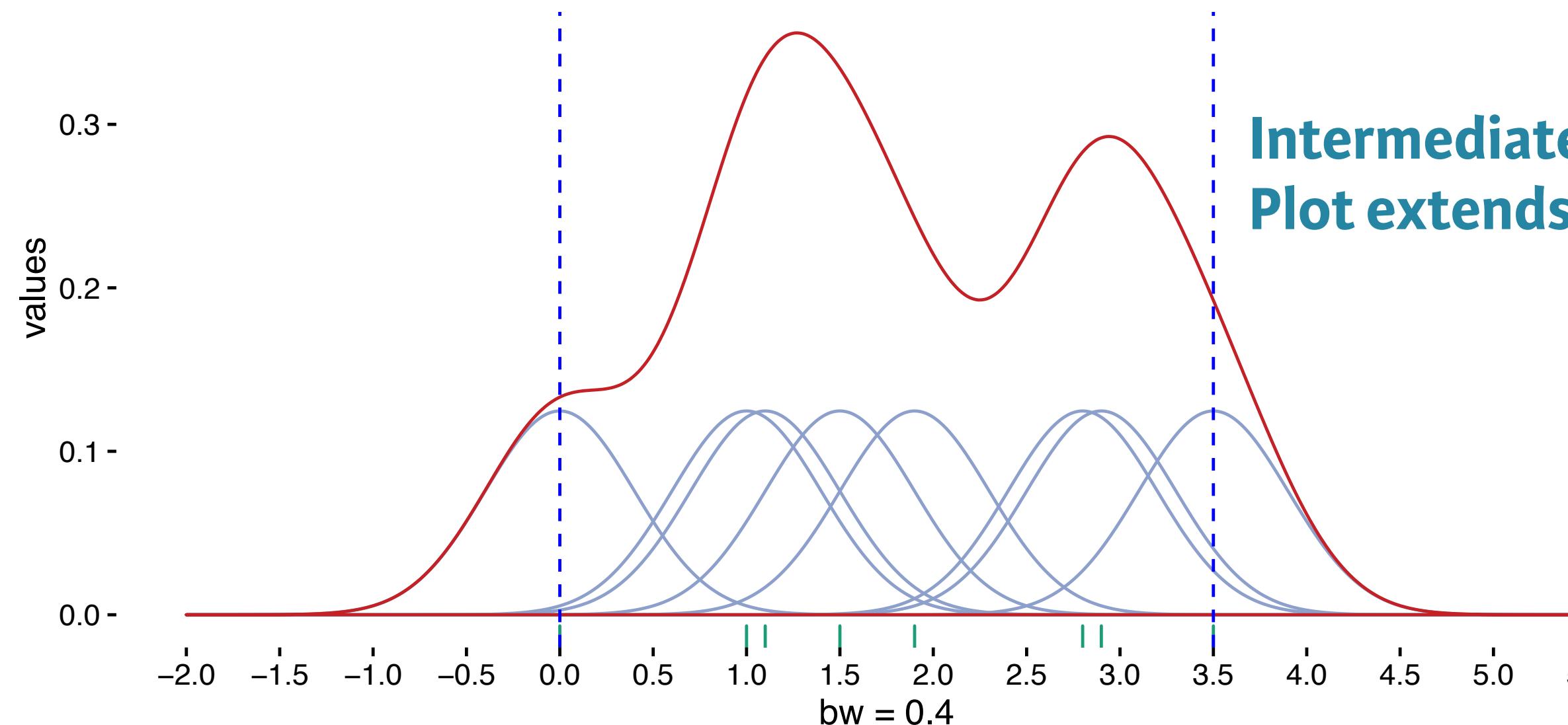


Remember:
Density plots are representations
of the underlying distribution!





geom_density()
happens for every bandwidth!





DATA VISUALIZATION WITH GGPLOT2

Let's practice!



DATA VISUALIZATION WITH GGPLOT2

Multiple Groups/Variables

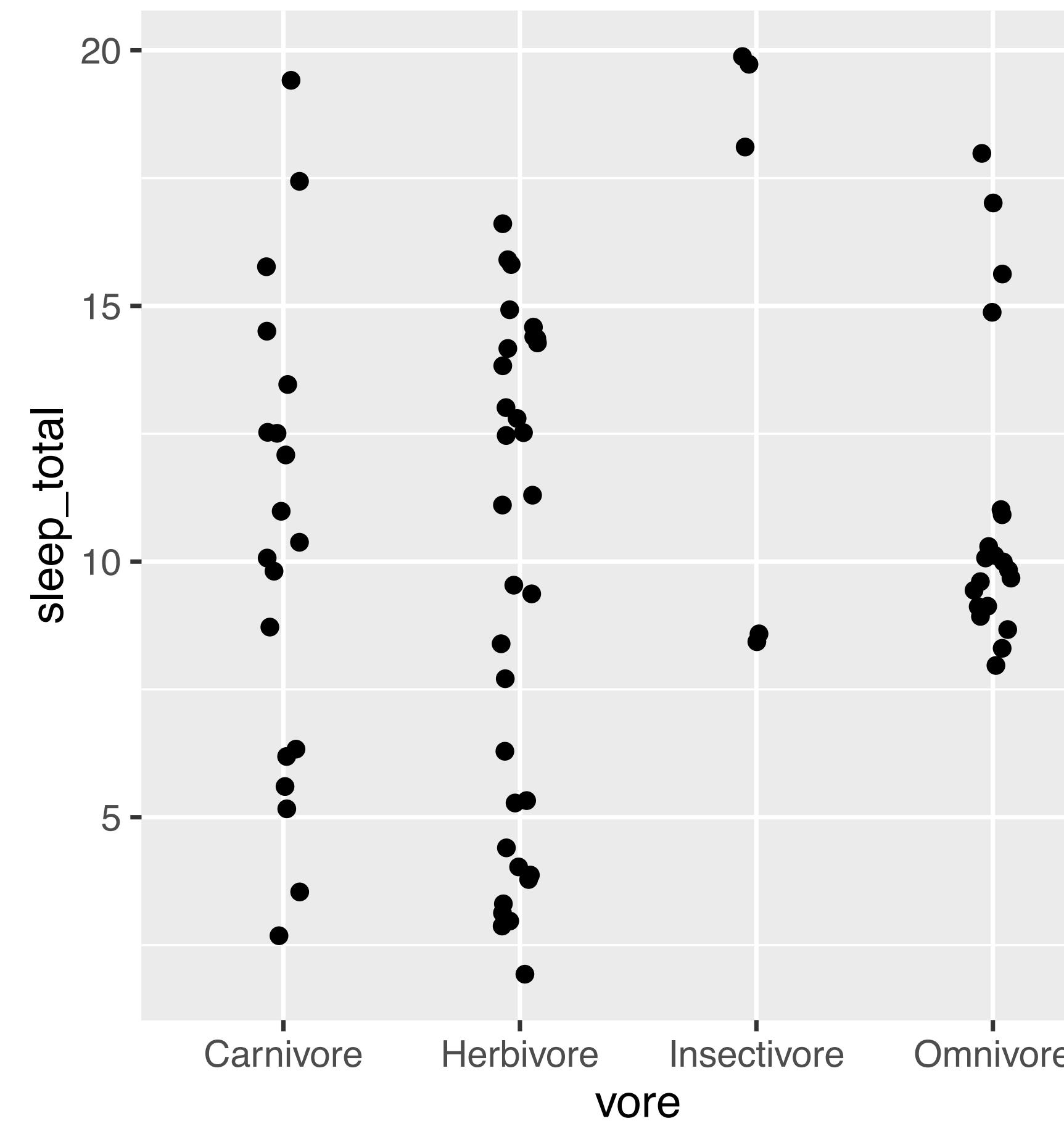
Groups

Levels within a factor variable

```
> head(mammals)
      vore sleep_total
1   Carnivore      12.1
2   Omnivore       17.0
3   Herbivore      14.4
4   Omnivore       14.9
5   Herbivore       4.0
6   Herbivore      14.4

> levels(mammals$vore)
[1] "Carnivore"    "Herbivore"    "Insectivore"  "Omnivore"
```

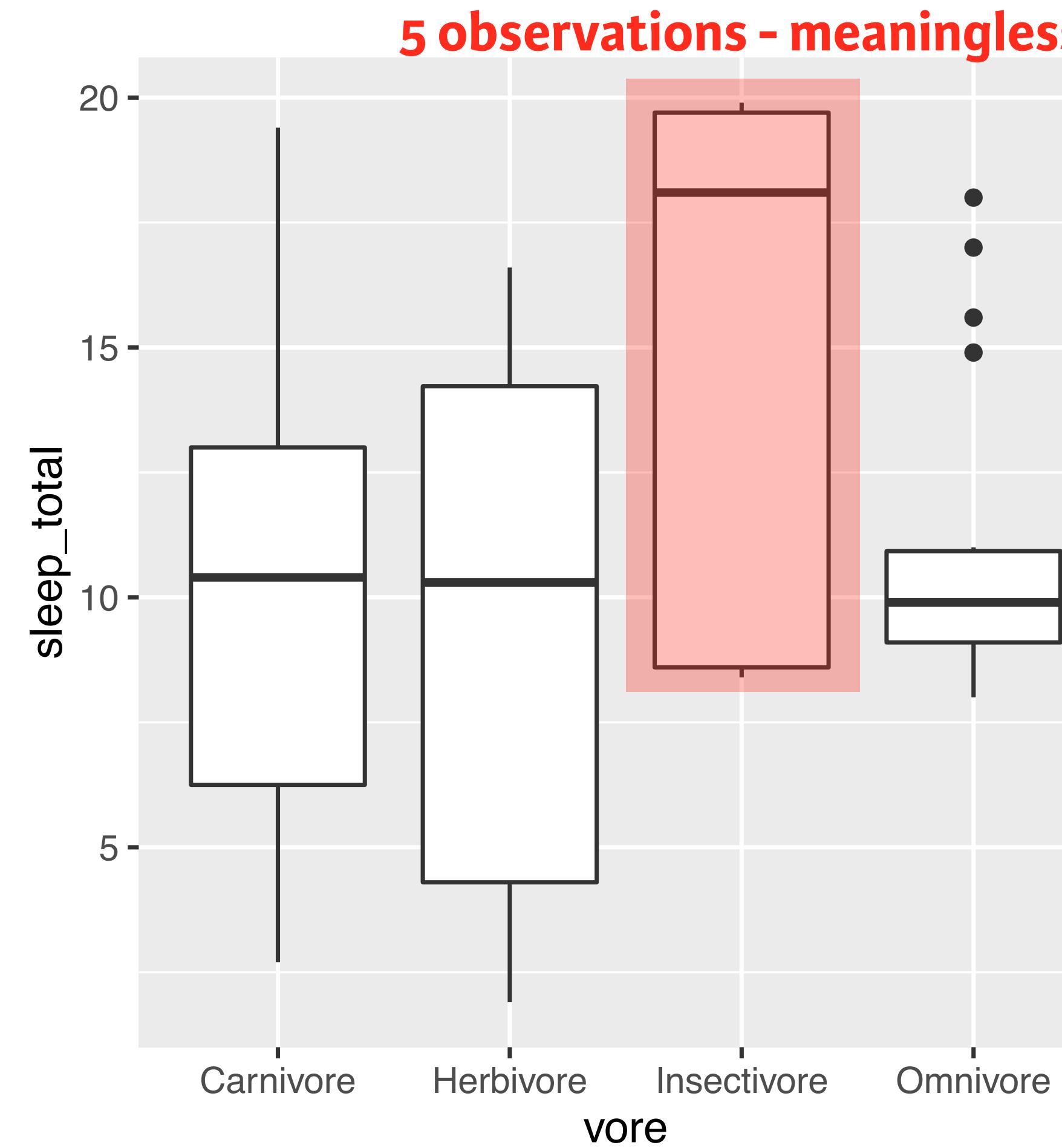
Jittered points



```
ggplot(mammals, aes(x = vore, y = sleep_total)) +  
  geom_point(position = position_jitter(0.2))
```

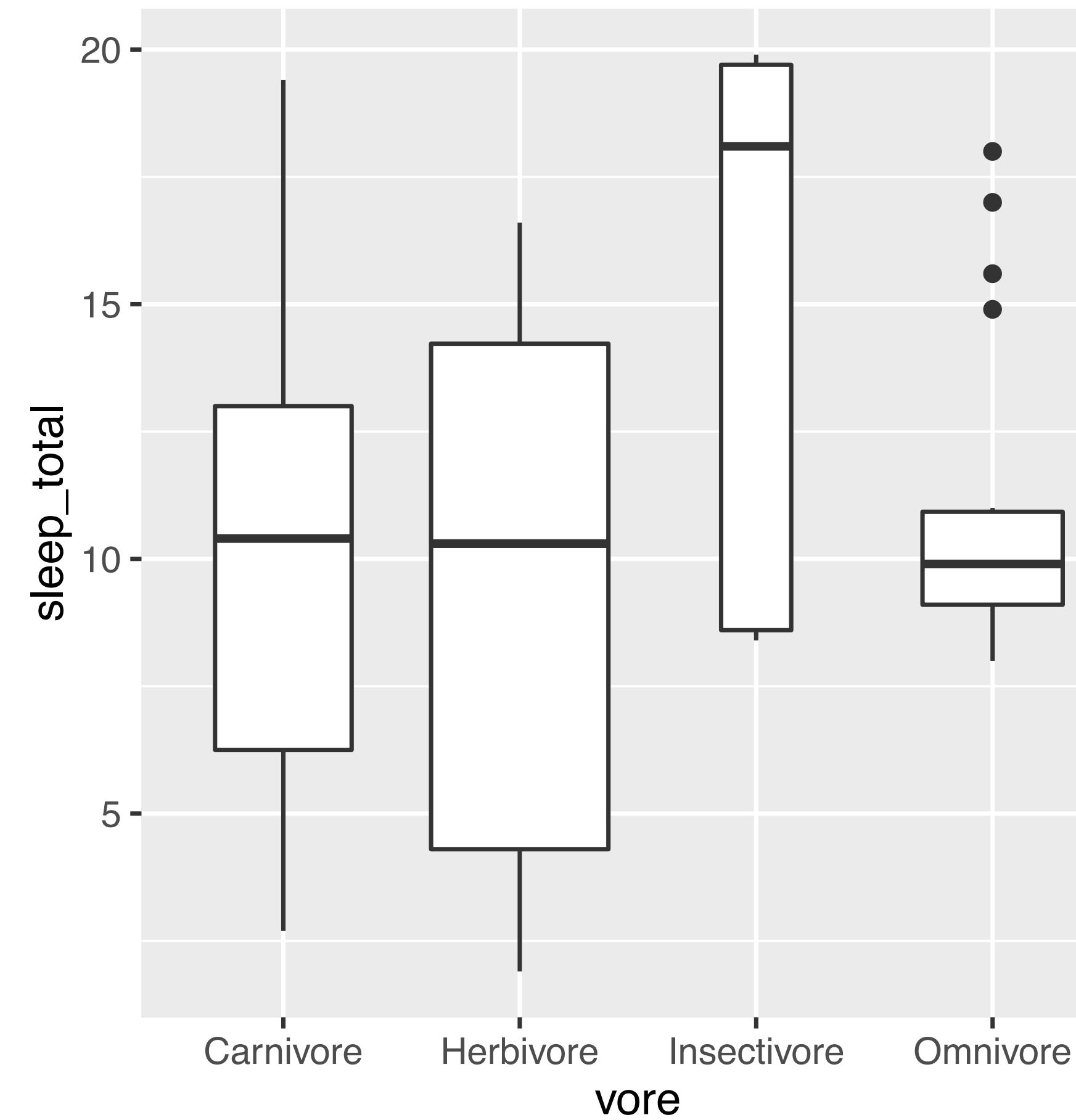
Box plot

```
ggplot(mammals, aes(x = vore, y = sleep_total)) +  
  geom_boxplot()
```

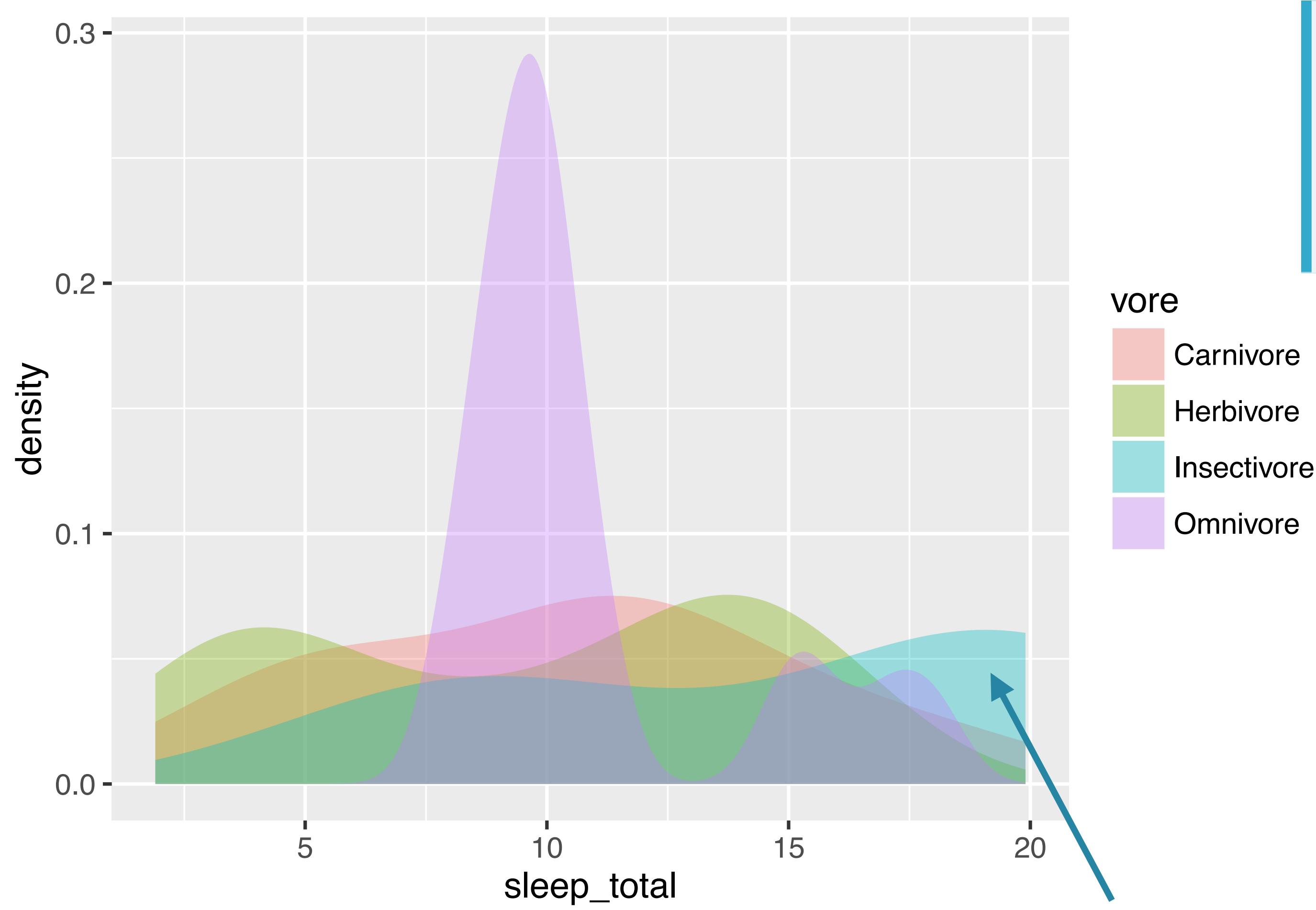


Box plot (2)

```
ggplot(mammals, aes(x = vore, y = sleep_total)) +  
  geom_boxplot(varwidth = TRUE)
```



Density plots



```
ggplot(mammals, aes(x = sleep_total, fill = vore)) +  
  geom_density(col = NA, alpha = 0.35)
```

```
> # Add weights  
> mammals <- mammals %>%  
  group_by(vore) %>%  
  mutate(n = n()/nrow(mammals))
```

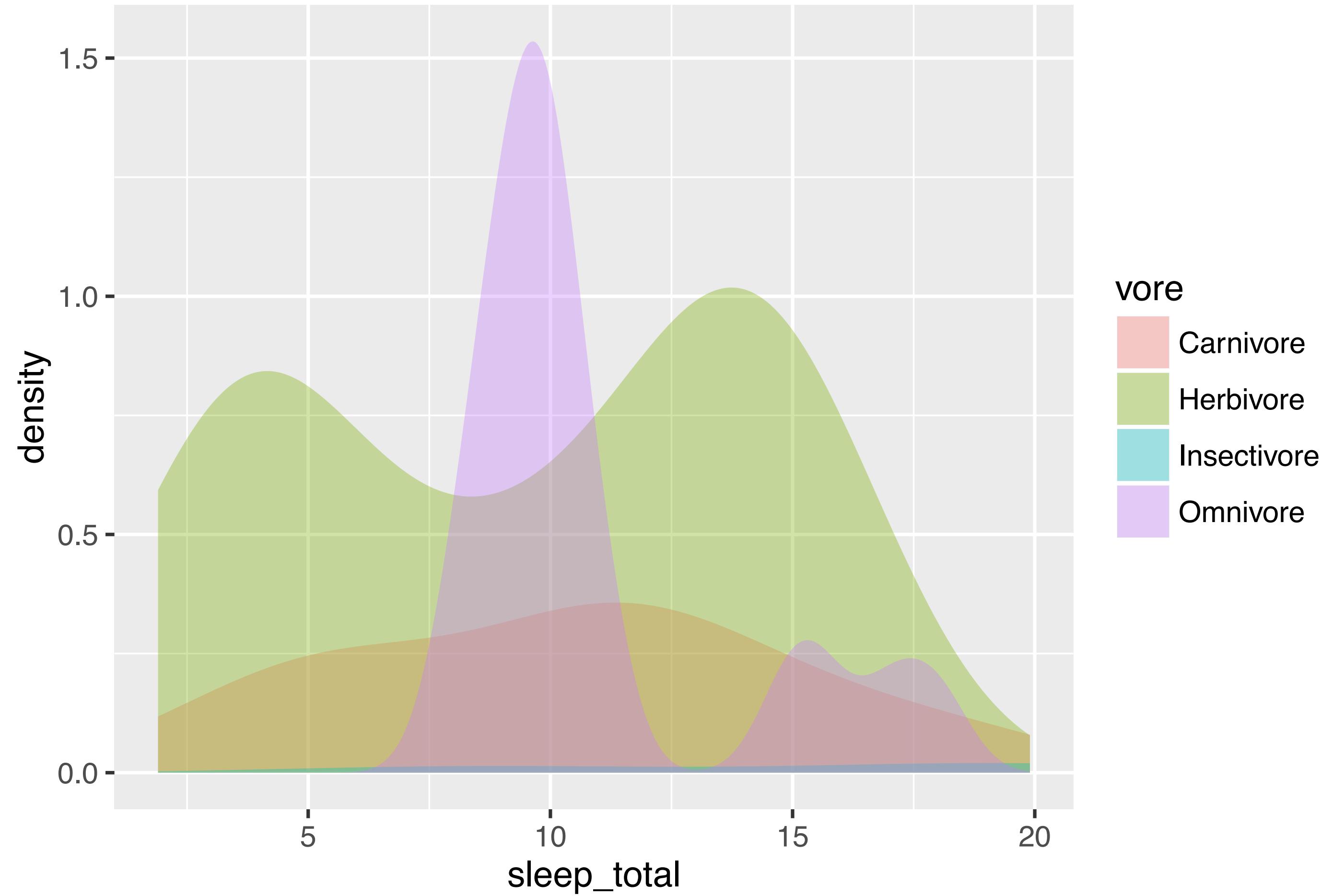
vore

- Carnivore
- Herbivore
- Insectivore
- Omnivore

abundant, but only 5 observations!

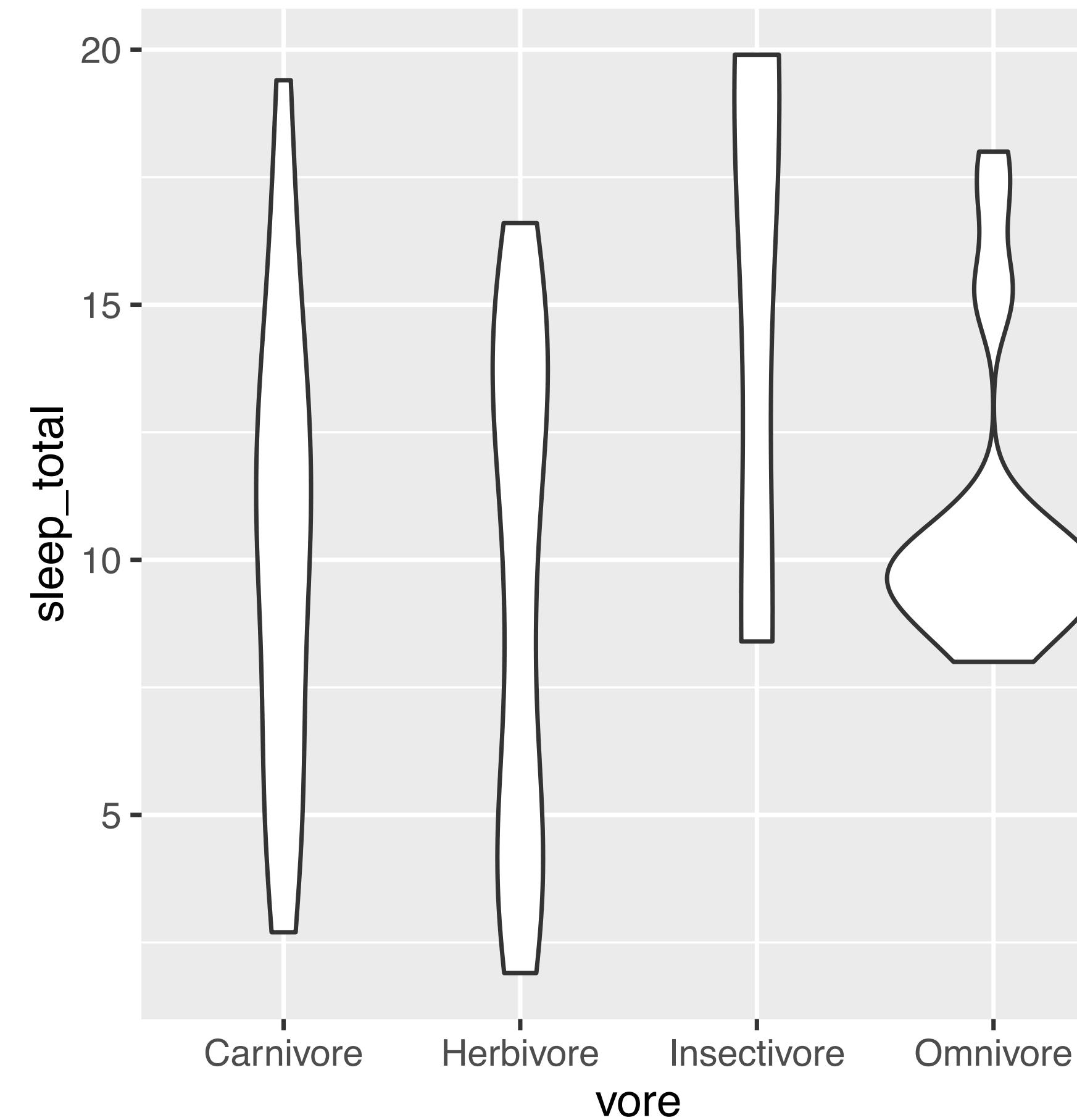
Weighted

```
ggplot(mammals, aes(x = sleep_total, fill = vore)) +  
  geom_density(aes(weight = n), col = NA, alpha = 0.35)
```



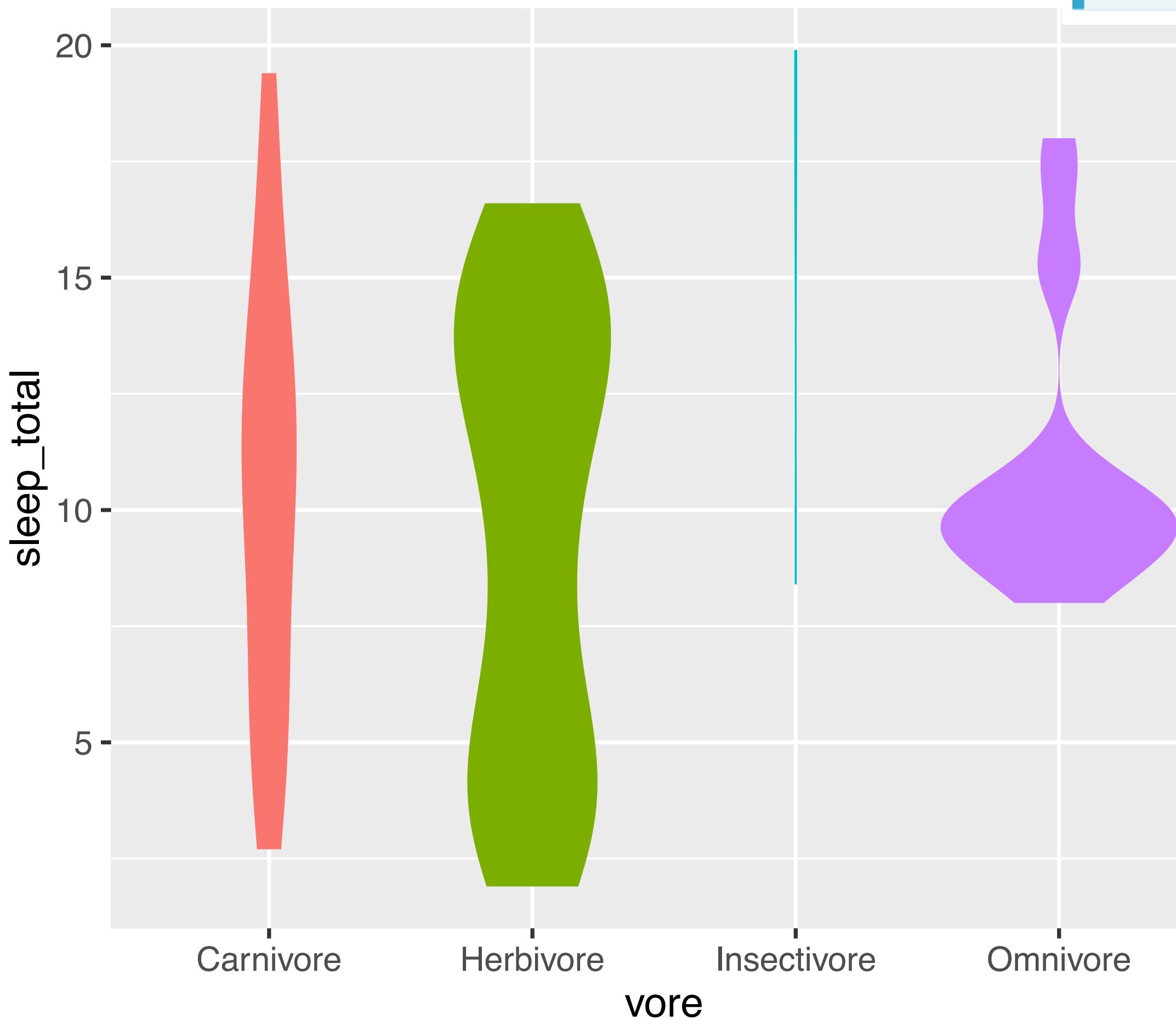
Violin plot

```
ggplot(mammals, aes(x = vore, y = sleep_total)) +  
  geom_violin()
```



Weighted

```
ggplot(mammals, aes(x = vore,  
                     y = sleep_total,  
                     fill = vore)) +  
  geom_violin(aes(weight = n), col = NA)
```



vore

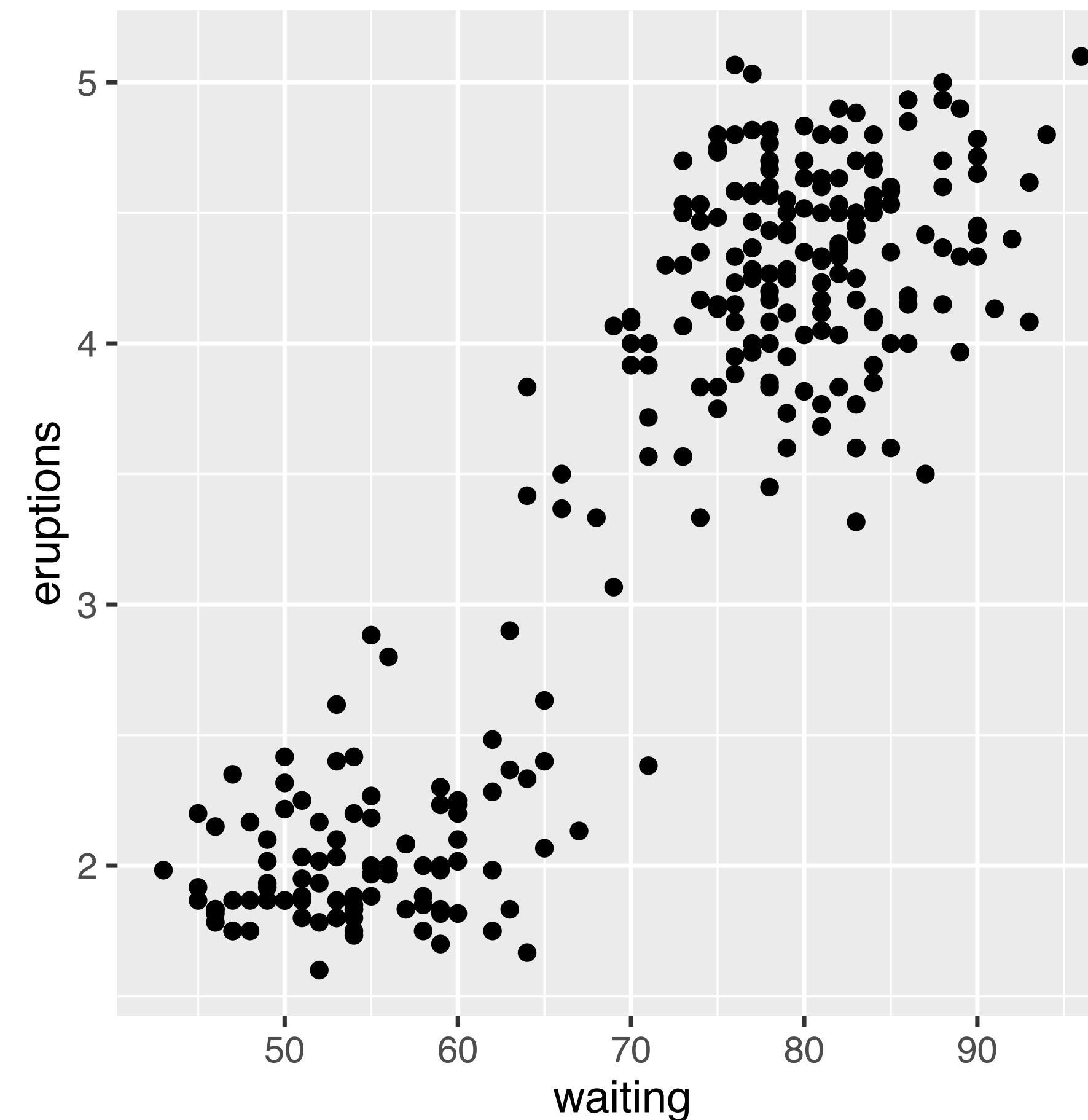
- Carnivore
- Herbivore
- Insectivore
- Omnivore

Compare separate variables

```
> dim(faithful)
[1] 272   2

> head(faithful)
  eruptions waiting
1      3.600      79
2      1.800      54
3      3.333      74
4      2.283      62
5      4.533      85
6      2.883      55
```

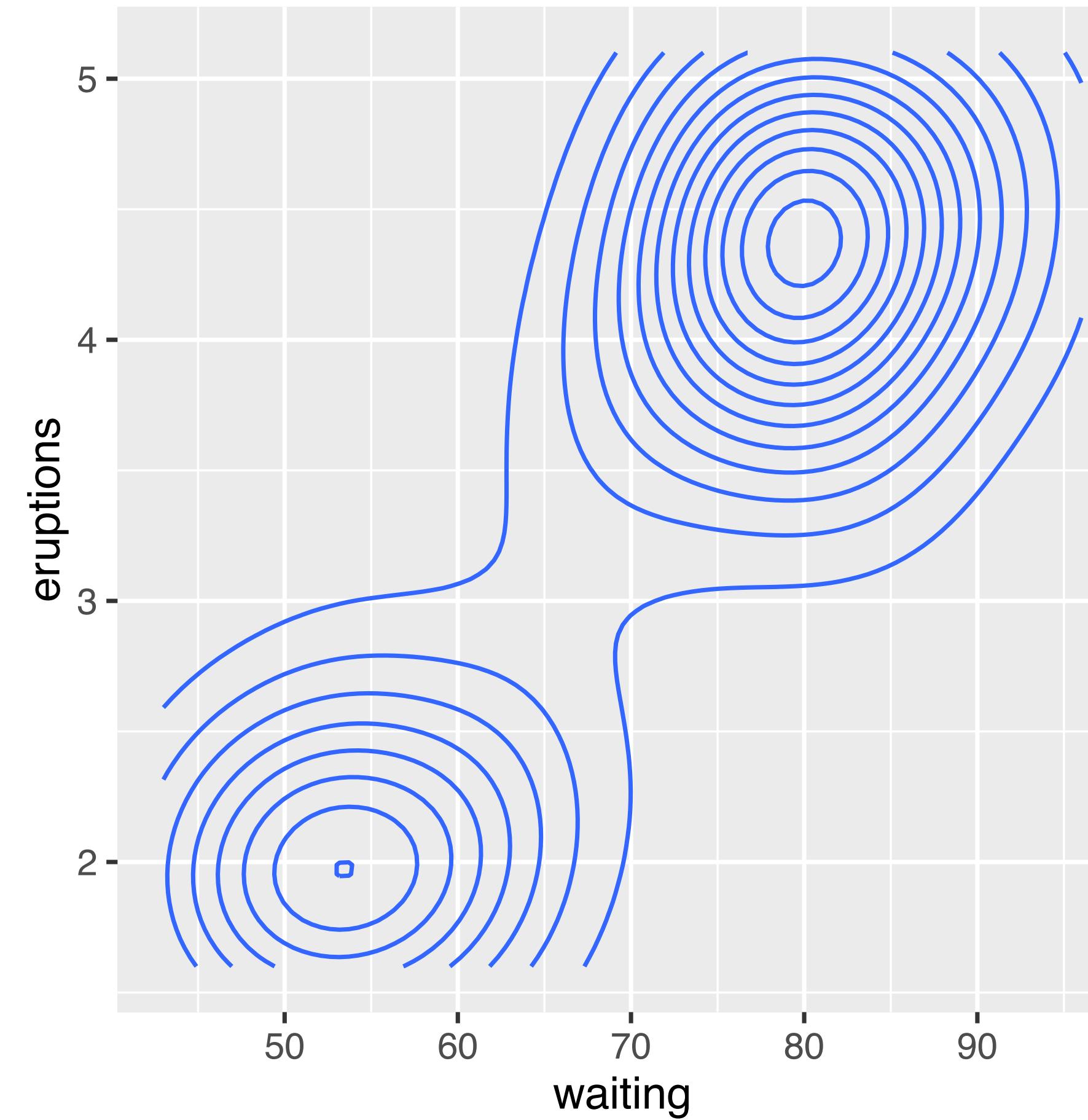
First look



```
ggplot(faithful, aes(x = waiting, y = eruptions)) +  
  geom_point()
```

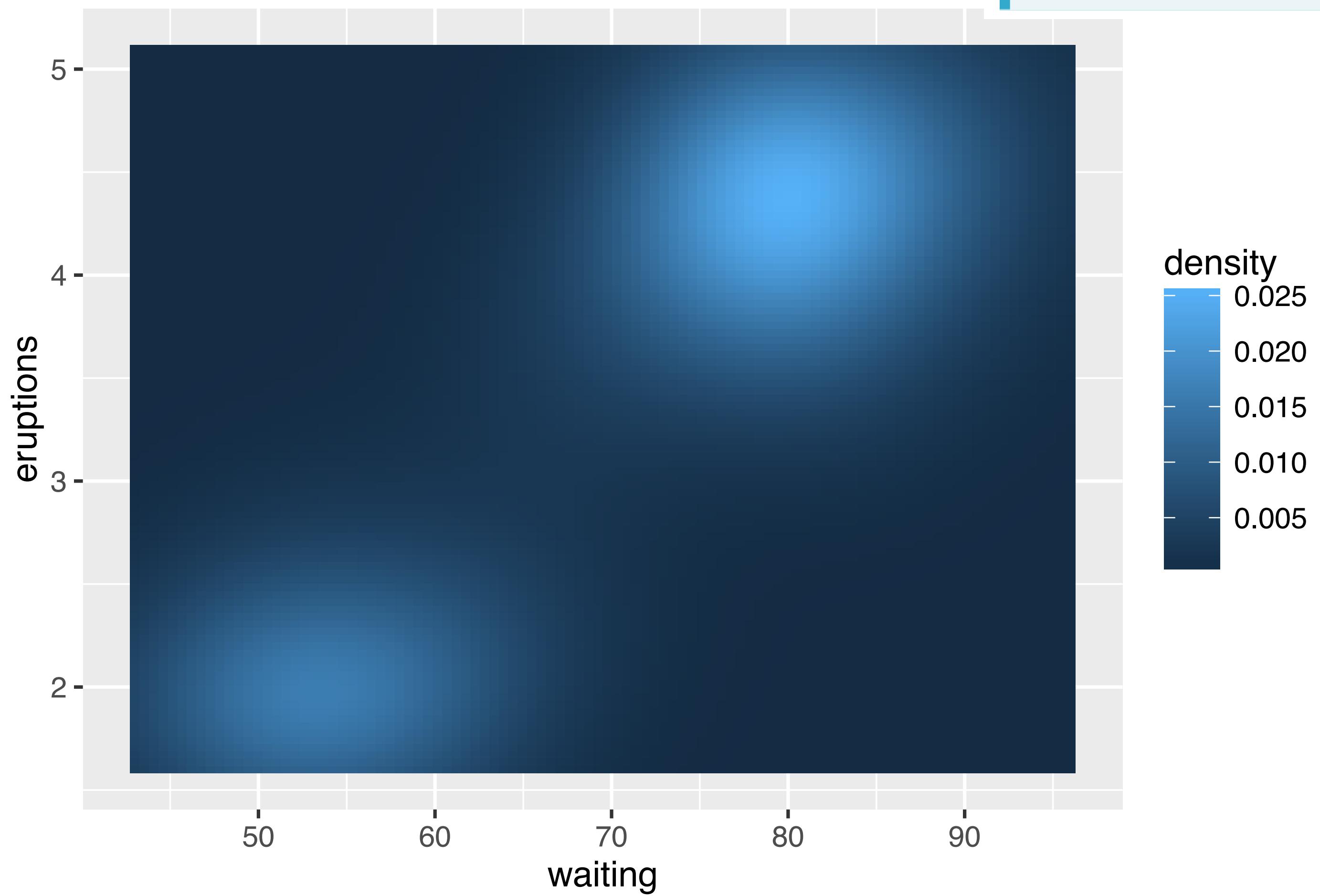
2D density plot

```
ggplot(faithful, aes(x = waiting, y = eruptions)) +  
  geom_density_2d()
```

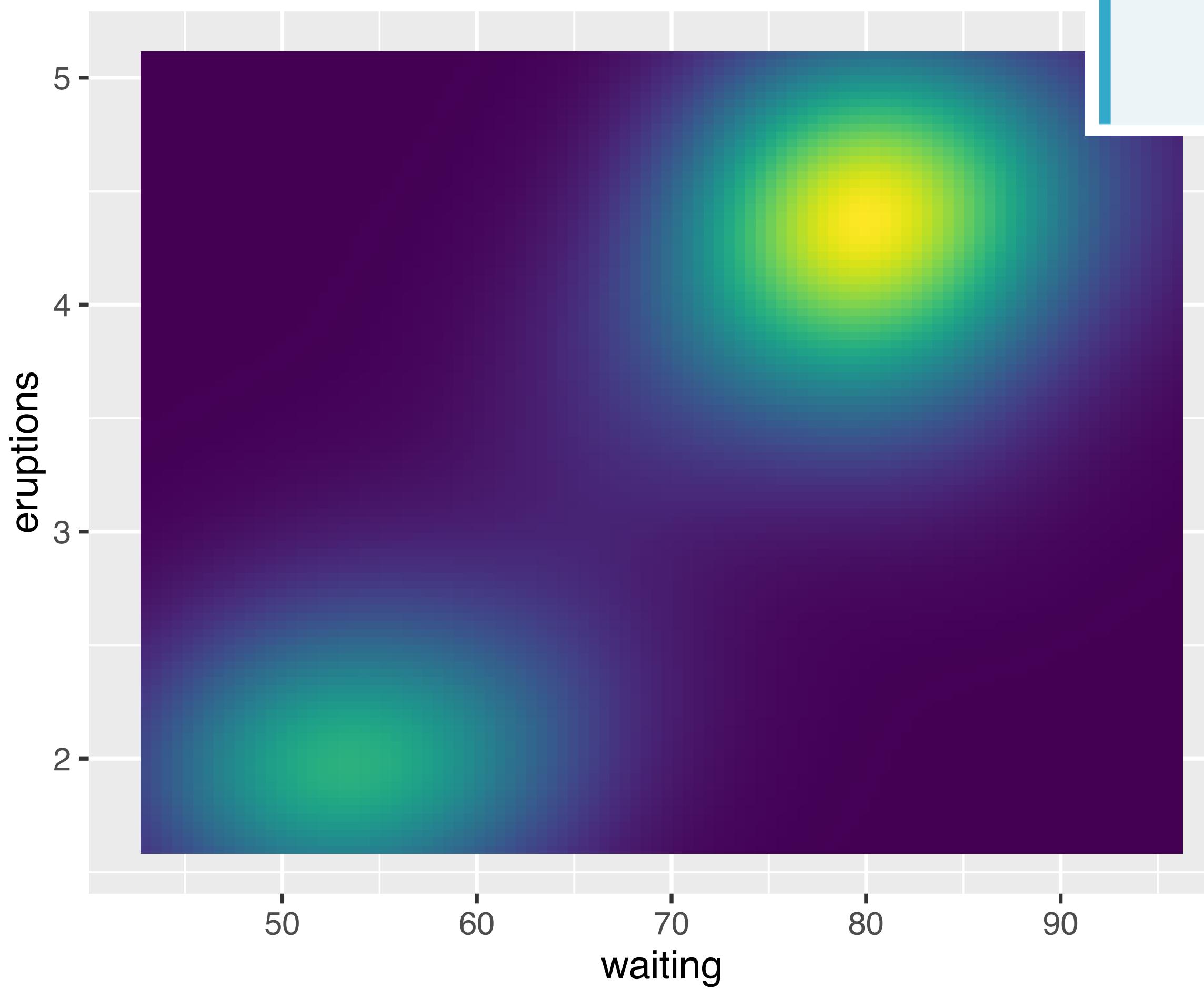


2D density plot

```
ggplot(faithful, aes(x = waiting, y = eruptions)) +  
  stat_density_2d(geom = "tile",  
    aes(fill = ..density..),  
    contour = FALSE)
```

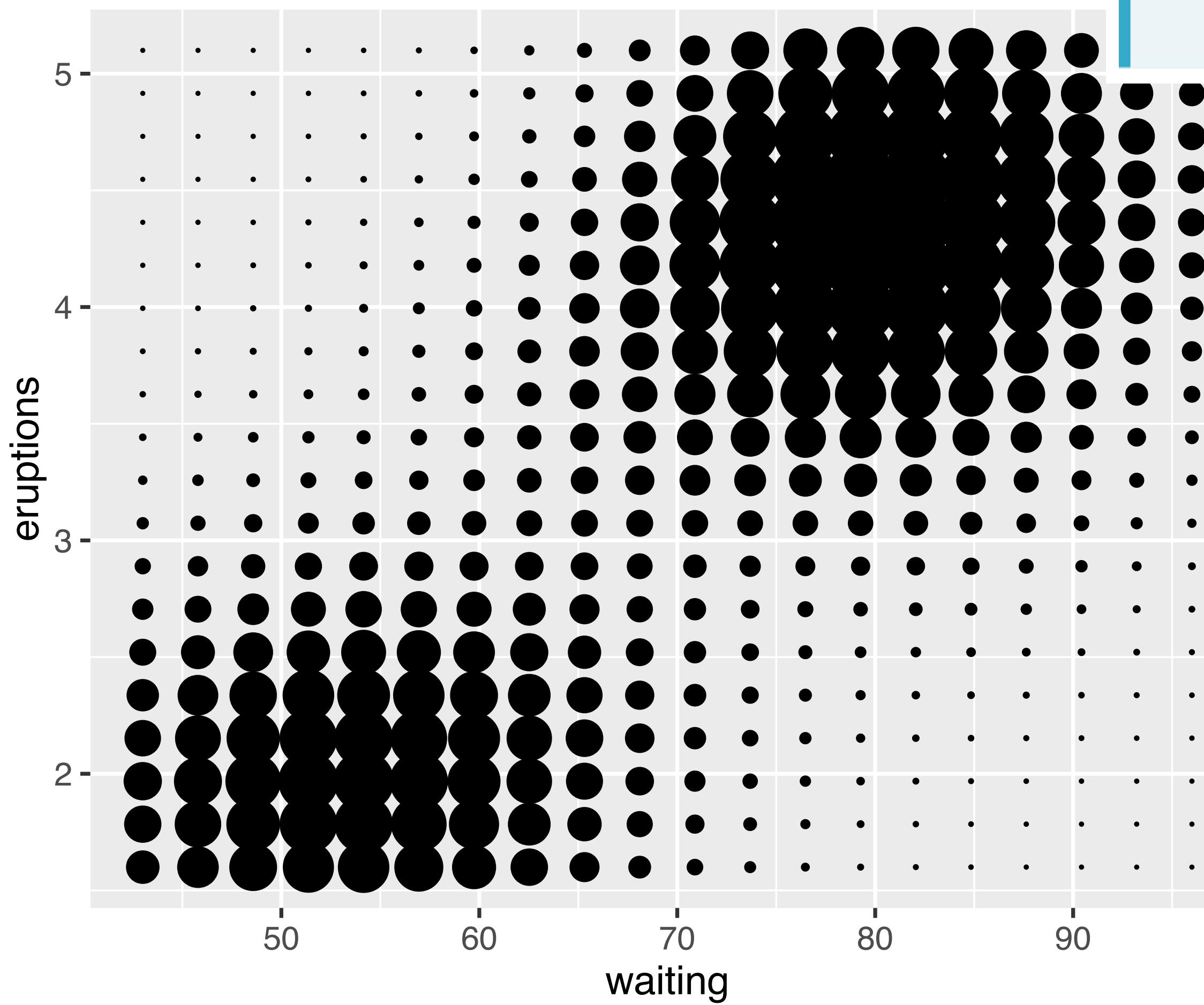


Viridis



```
library(viridis)
ggplot(faithful, aes(x = waiting, y = eruptions)) +
  stat_density_2d(geom = "tile",
                  aes(fill = ..density..),
                  contour = FALSE) +
  scale_fill_viridis()
```

Grid of circles



```
ggplot(faithful, aes(x = waiting, y = eruptions)) +  
  stat_density_2d(geom = "point",  
    aes(size = ..density..),  
    n = 20, contour = FALSE) +  
  scale_size(range = c(0, 9))
```



DATA VISUALIZATION WITH GGPLOT2

Let's practice!