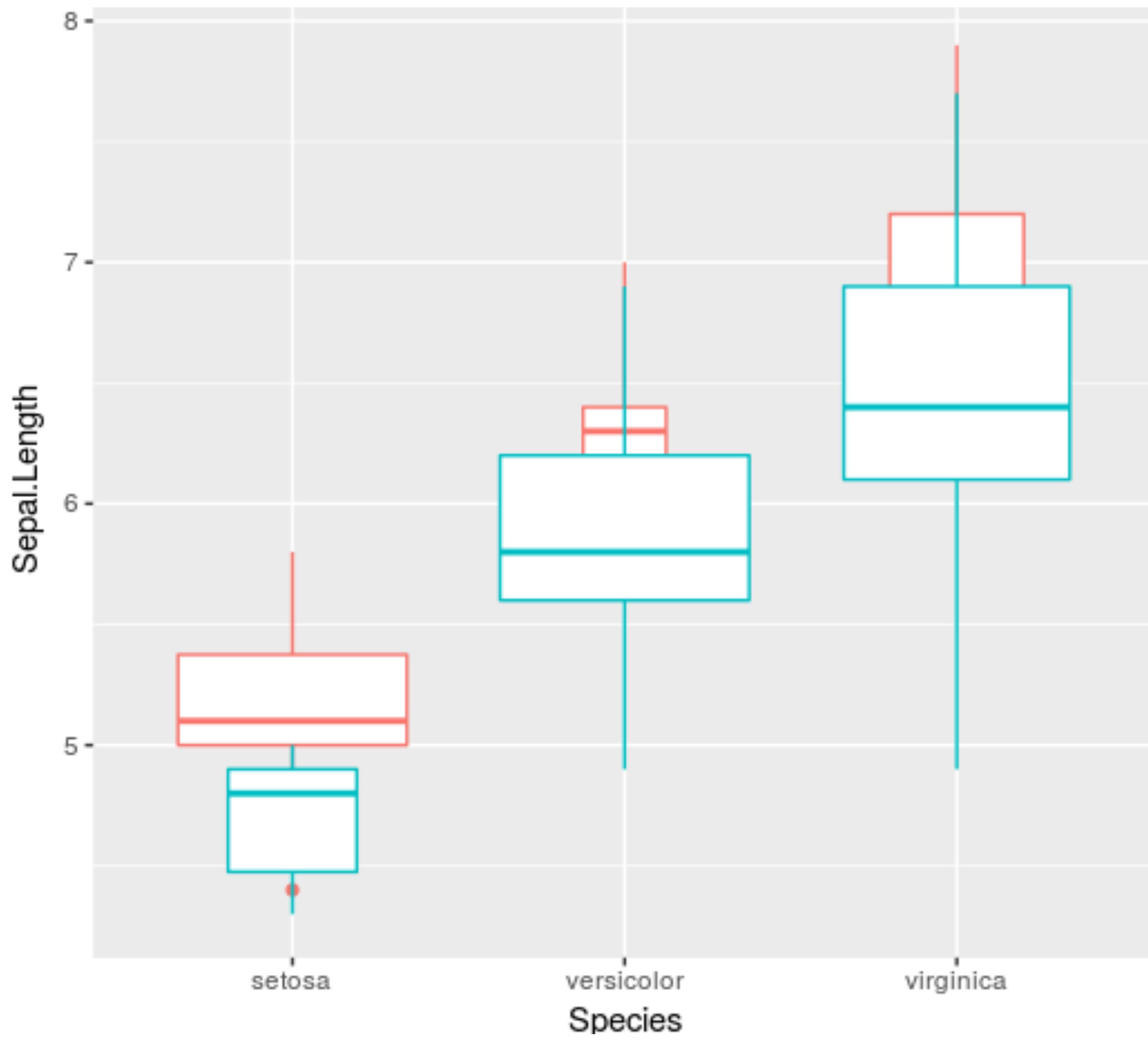


BOX PLOTS

A CASE STUDY IN DEBUGGING AND PERSEVERANCE

Kara Woo | @kara_woo
Sage Bionetworks
rstudio::conf 2019

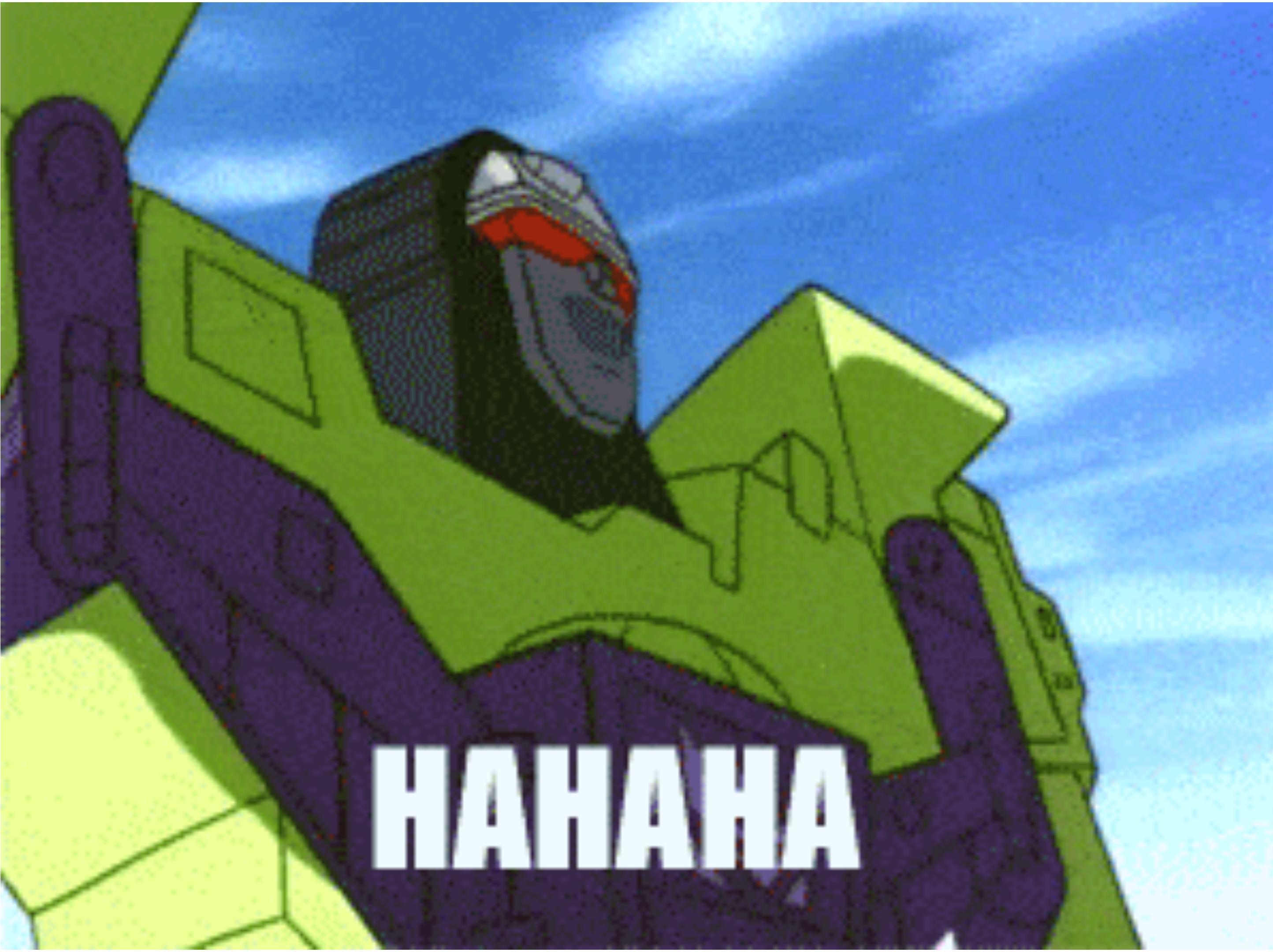
THIS IS ONLY KIND OF
ABOUT BOX PLOTS



"When I try to produce boxplots with colours depending on a categorical variable, these appear overlapping if varwidth is set to TRUE"

—GitHub user mcol

“This should be straightforward.”



How do I know what the bug is?

How do I fix it?

How do I know when I'm done?

HOW DO YOU KNOW WHAT THE BUG IS?

Isolate the
problem



Photo: Jia Ye

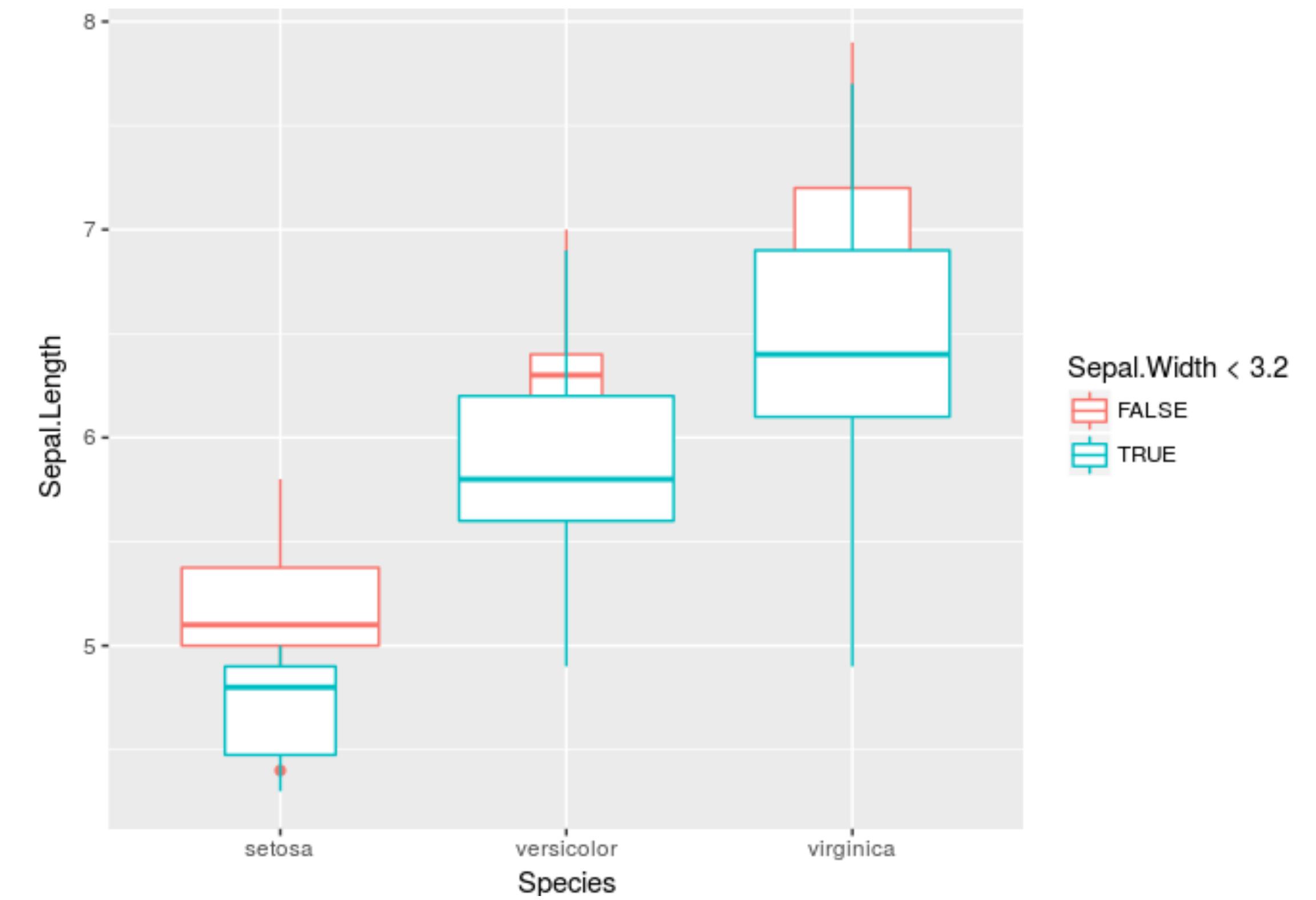


REPREX

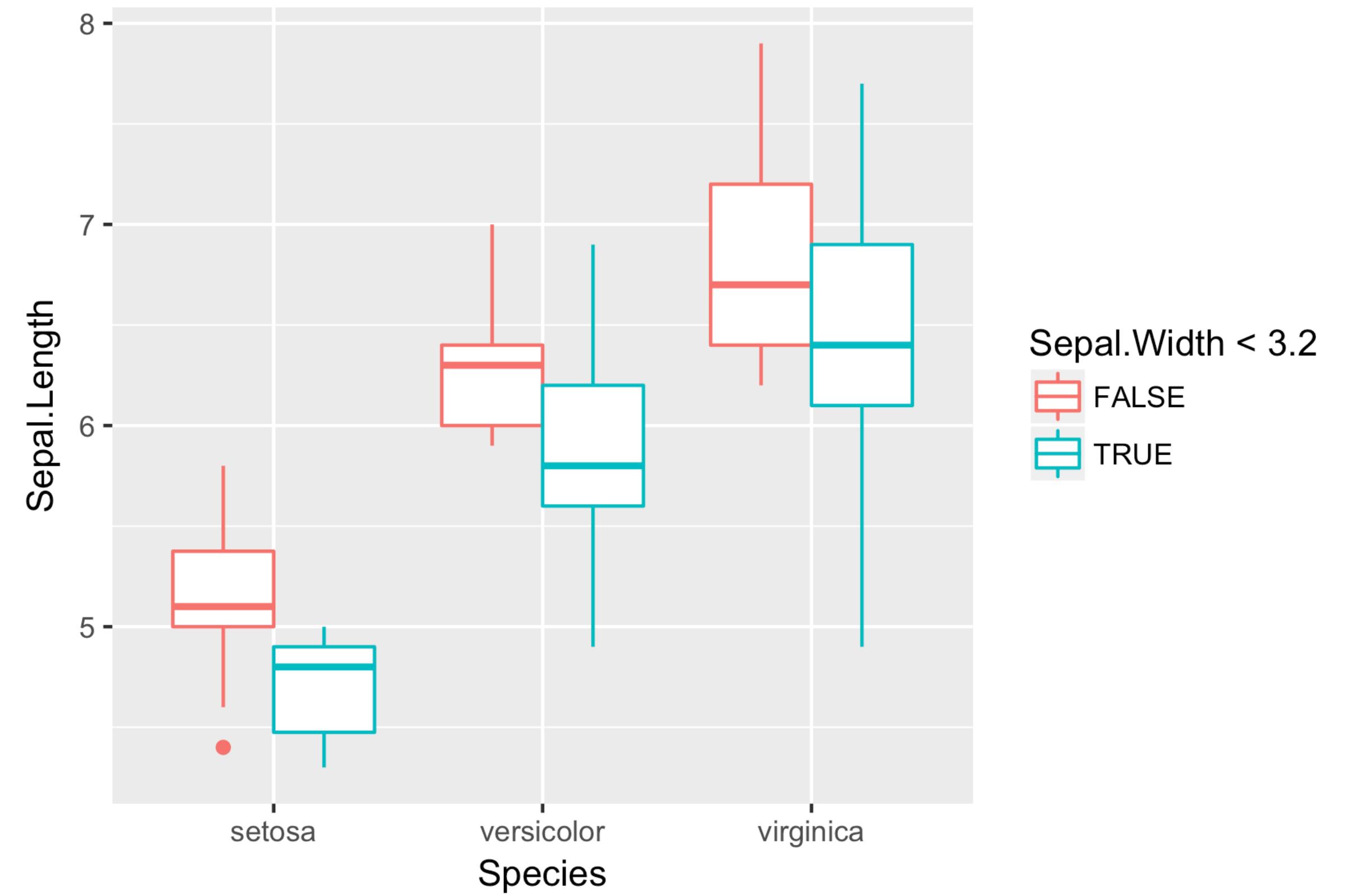
minimal **reproducible example**

`reprex::reprex()`

```
require(ggplot2)
#> Loading required package: ggplot2
ggplot(data = iris, aes(Species, Sepal.Length)) +
  geom_boxplot(aes(colour = Sepal.Width < 3.2), varwidth = TRUE)
#> Warning: position_dodge requires non-overlapping x intervals
```



```
require(ggplot2)
#> Loading required package: ggplot2
ggplot(data = iris, aes(Species, Sepal.Length)) +
  geom_boxplot(aes(colour = Sepal.Width < 3.2), varwidth = FALSE)
```

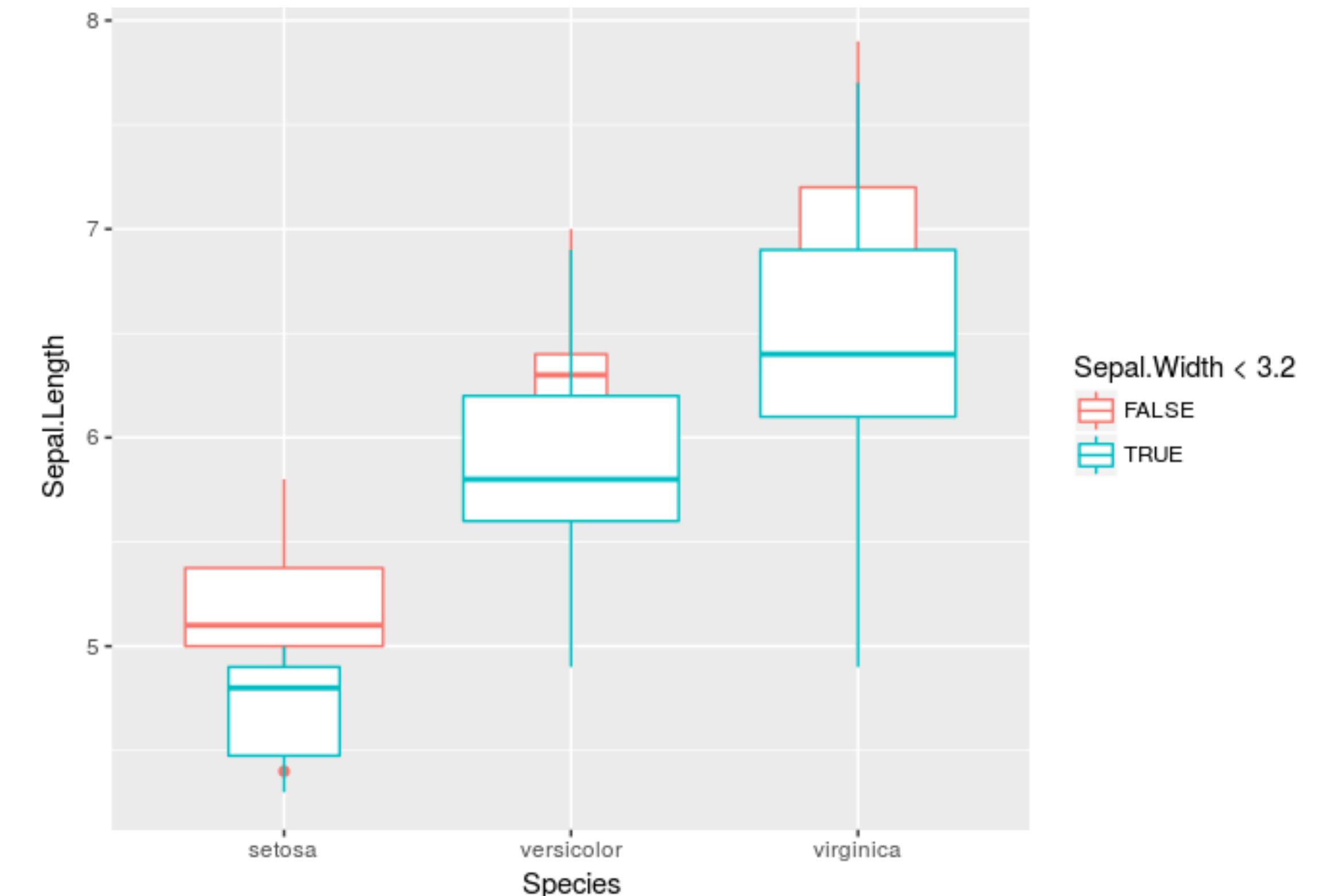


Follow
the trails



Photo: Jia Ye

```
require(ggplot2)
#> Loading required package: ggplot2
ggplot(data = iris, aes(Species, Sepal.Length)) +
  geom_boxplot(aes(colour = Sepal.Width < 3.2), varwidth = TRUE)
#> Warning: position_dodge requires non-overlapping x intervals
```



Search · position_dodge

GitHub, Inc. [US] | https://github.com/tidyverse/ggplot2/search?q=position_dodge&unscoped_q=position_dodge

position_dodge

Pull requests Issues Marketplace Explore

Code 14

Commits 10

Issues 59

Wikis

Languages

R 11

Markdown 1

Text 1

14 code results in [tidyverse/ggplot2](#) or view all results on GitHub

Sort: Best match ▾

R/position-dodge.r R

Showing the top two matches Last indexed on Jul 17, 2018

```
3 #' Dodging preserves the vertical position of an geom while adjusting the
4 #' horizontal position. `position_dodge2` is a special case of `position_dodge`
...
22 #' ggplot(mtcars, aes(factor(cyl), fill = factor(vs))) +
23 #'   geom_bar(position = position_dodge(preserve = "total"))
24 #
25 #' \donttest{
```

tests/testthat/test-position_dodge.R R

Showing the top two matches Last indexed on Nov 15, 2018

```
1 context("position_dodge")
2
3 test_that("can control whether to preserve total or individual width", {
...
6 p_total <- ggplot(df, aes(x, fill = y)) +
7   geom_bar(position = position_dodge(preserve = "total"), width = 1)
8 p_single <- ggplot(df, aes(x, fill = y)) +
```

R/geom-text.r R

Showing the top two matches Last indexed on Sep 17, 2018

```
99 #' geom_text(aes(label = y), position = position_dodge(0.9))
100 #' # Use you can't nudge and dodge text, so instead adjust the y position
...
102 #' geom_col(aes(fill = grp), position = "dodge") +
103 #'   geom_text(
104 #'     aes(label = v, v = v + 0.05).
```

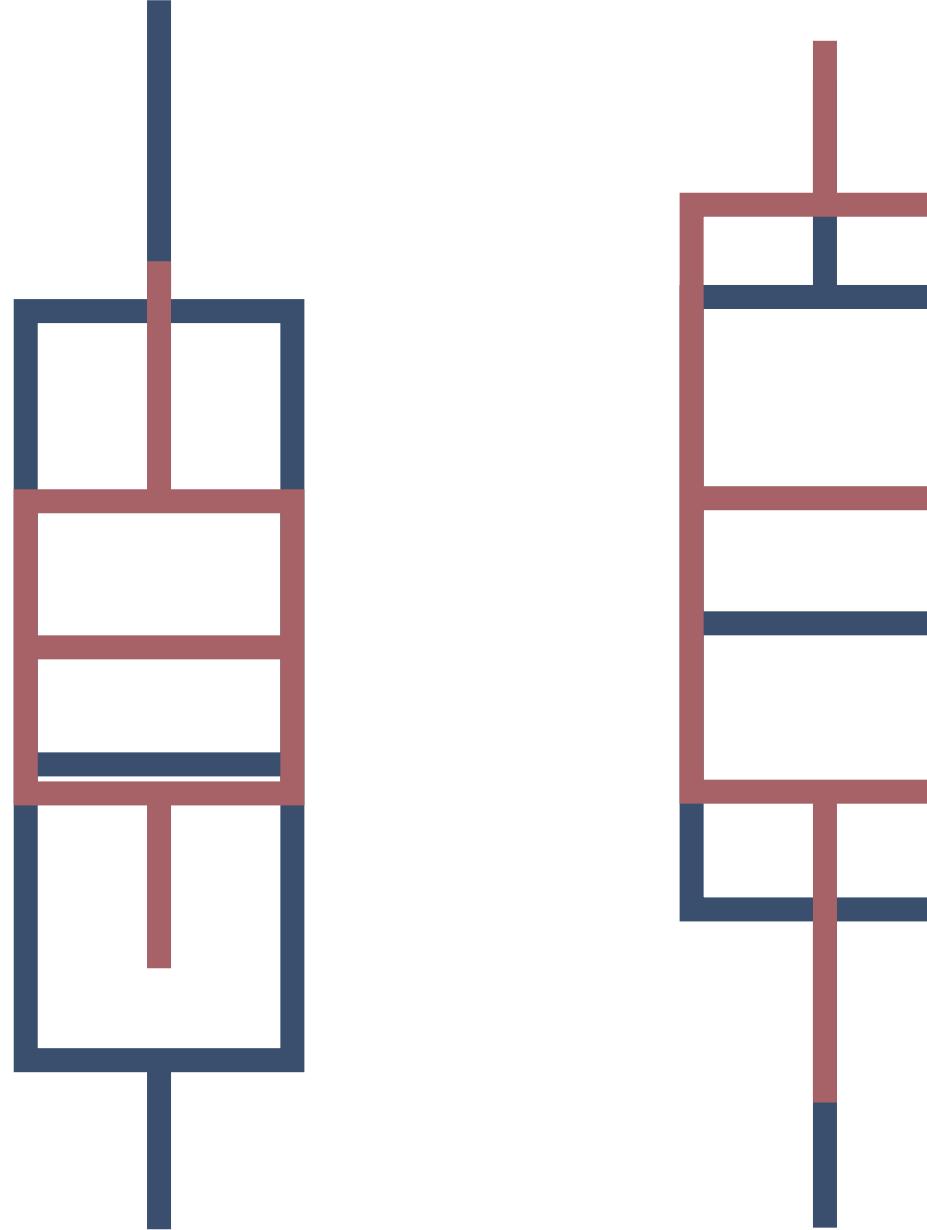
```
59     #' position = position_dodge(width = 0.9)
60   #' )
61   position_dodge <- function(width = NULL, preserve = "all") {
62     ggproto(NULL, PositionDodge,
63       width = width,
64       preserve = preserve)
65   }
66 }
67
1 # Detect and prevent collisions
2 # Powers dodging, stacking
3 collide <- function(data,
4   # Determine width
5   if (!is.null(width)) {
6     # Width set manually
7     if (!all(c("xmin", "xmax") %in% names(data))) {
8       data$xmin <- data$x - width / 2
9       data$xmax <- data$x + width / 2
10    }
11  } else {
12    if (!all(c("xmin", "xmax") %in% names(data))) {
```

COLLIDE()

Gets information about box location

Looks for box overlap

Passes boxes that share position to
pos_dodge()



POS_DODGE()

Scales boxes down

Places boxes side by side

```
> debug(ggplot2:::collide)
```

```
> debug(ggplot2:::collide)
> ggplot(data = iris, aes(Species, Sepal.Length)) +
>   geom_boxplot(aes(colour = Sepal.Width < 3.2), varwidth = FALSE)
#> debugging in: collide(data, params$width,
#> name = "position_dodge", strategy = pos_dodge,
#> n = params$n, check.width = FALSE)
Browse[2]>
```

```
> debug(ggplot2:::collide)
> ggplot(data = iris, aes(Species, Sepal.Length)) +
>   geom_boxplot(aes(colour = Sepal.Width < 3.2), varwidth = FALSE)
#> debugging in: collide(data, params$width,
#> name = "position_dodge", strategy = pos_dodge,
#> n = params$n, check.width = FALSE)
```

Browse[2]> **data**

```
#> ... x xmin xmax
#> 1 ... 1 0.625 1.375
#> 2 ... 2 1.625 2.375
#> 3 ... 3 2.625 3.375
#> 4 ... 1 0.625 1.375
#> 5 ... 2 1.625 2.375
#> 6 ... 3 2.625 3.375
```

VARWIDTH = FALSE

data

```
#> ... x xmin xmax
#> 1 ... 1 0.625 1.375
#> 2 ... 2 1.625 2.375
#> 3 ... 3 2.625 3.375
#> 4 ... 1 0.625 1.375
#> 5 ... 2 1.625 2.375
#> 6 ... 3 2.625 3.375
```

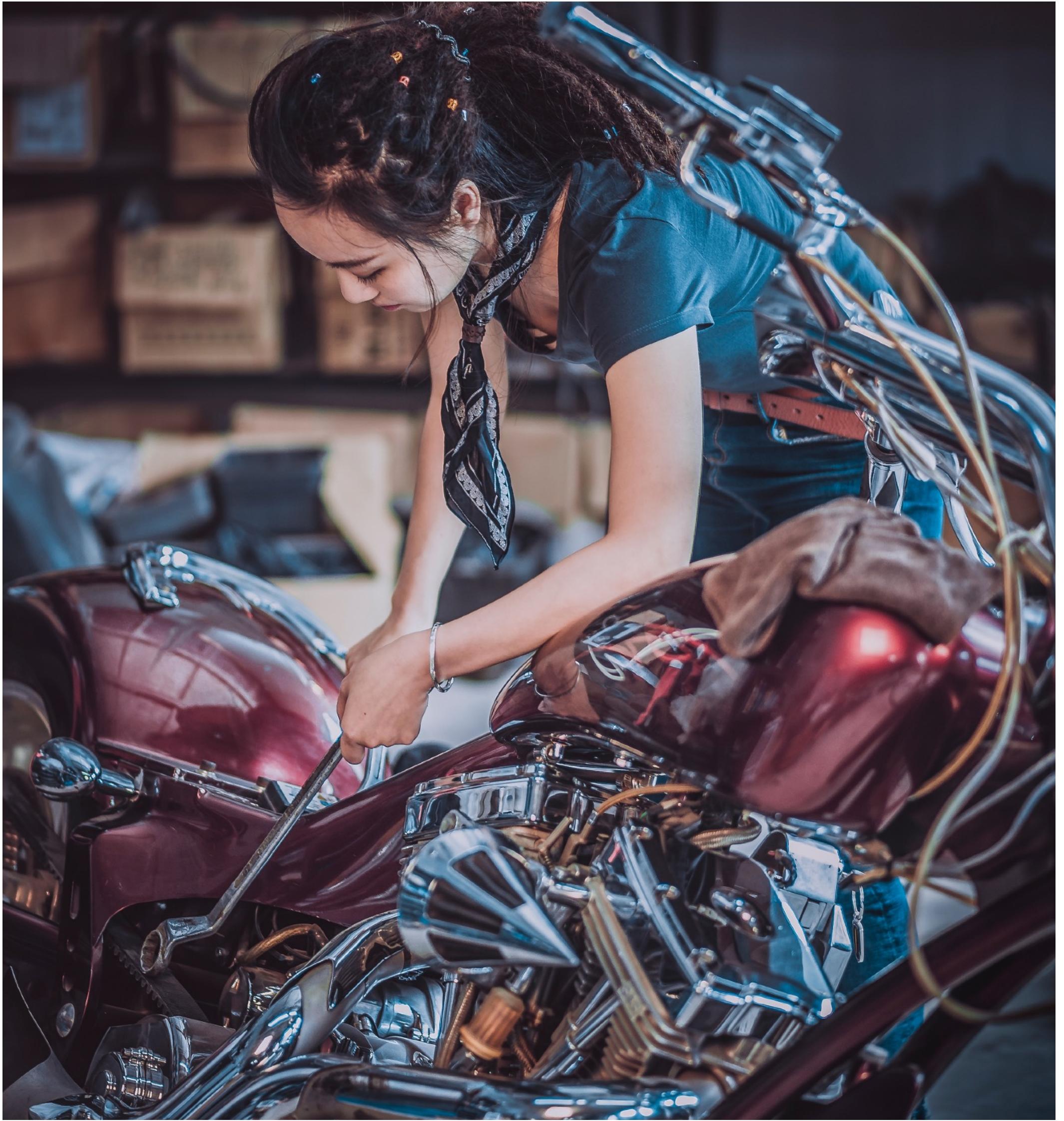
VARWIDTH = TRUE

data

```
#> ... x xmin xmax
#> 1 ... 1 0.6553988 1.344601
#> 2 ... 2 1.8750000 2.125000
#> 3 ... 3 2.7984436 3.201556
#> 4 ... 1 0.8063508 1.193649
#> 5 ... 2 1.6250000 2.375000
#> 6 ... 3 2.6599632 3.340037
```

```
collide <- function(data, ...) {  
  # ...  
  plyr::ddply(data, "xmin", strategy, ..., width = width)  
  # ...  
}
```

Boxes with different xmin aren't
treated as the same position.



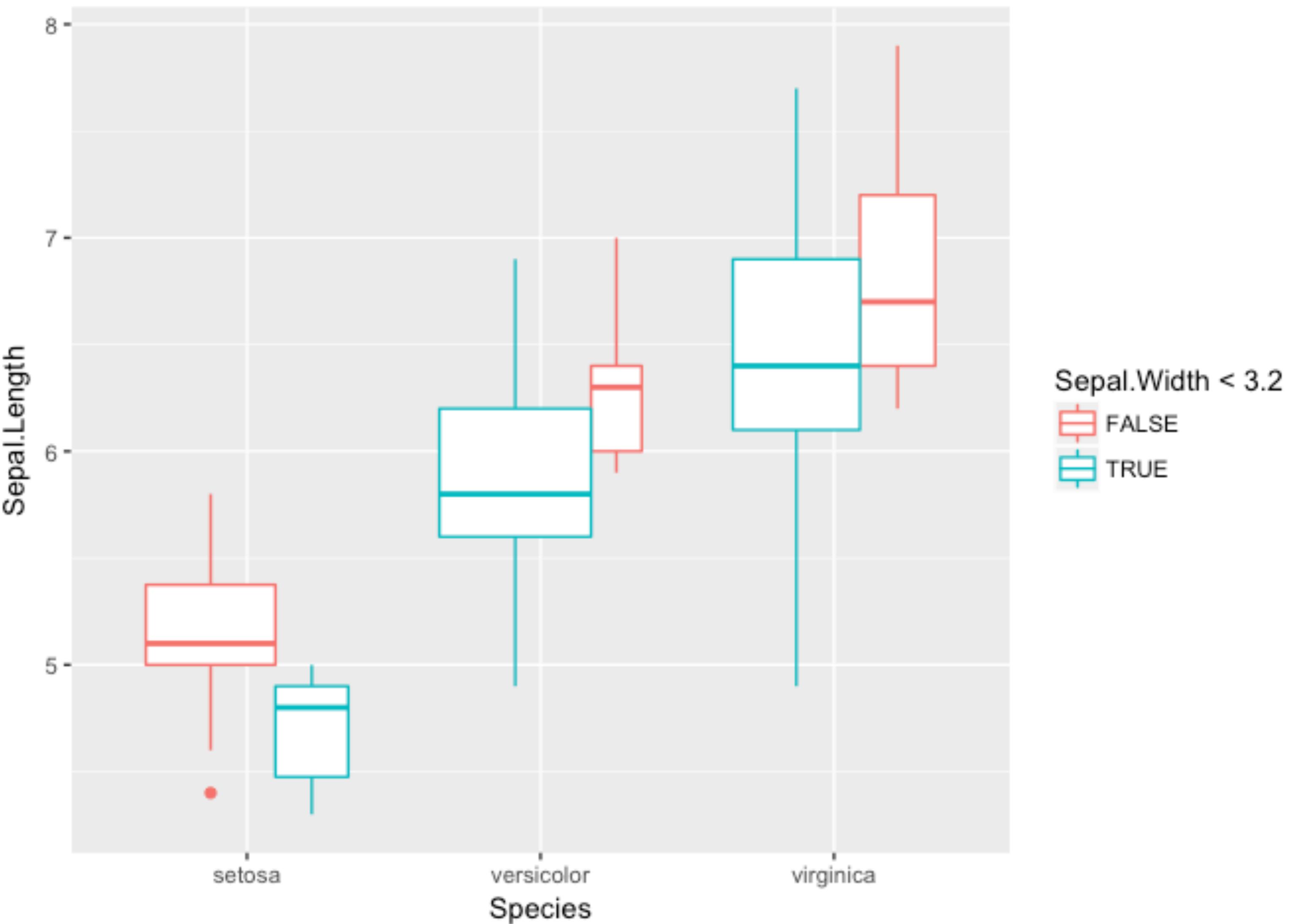
HOW DO
YOU FIX IT?

Experiment

- ccc6bbb4 This doesn't fix position_dodge, but it might be in the right direction?
- f5946680 Commit before I break something else

INITIAL “FIX”

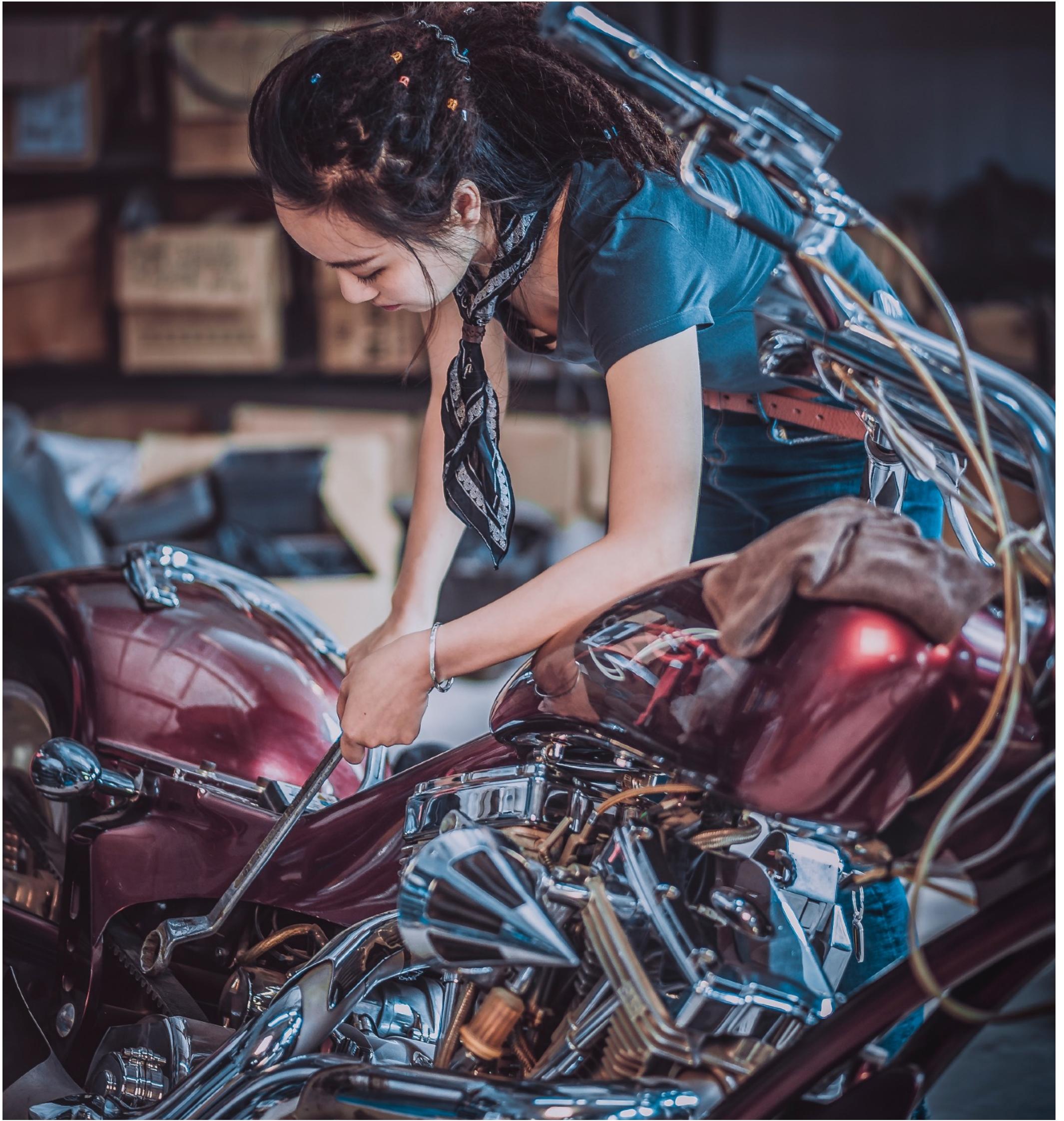
What's wrong with
this picture?



INITIAL “FIX”

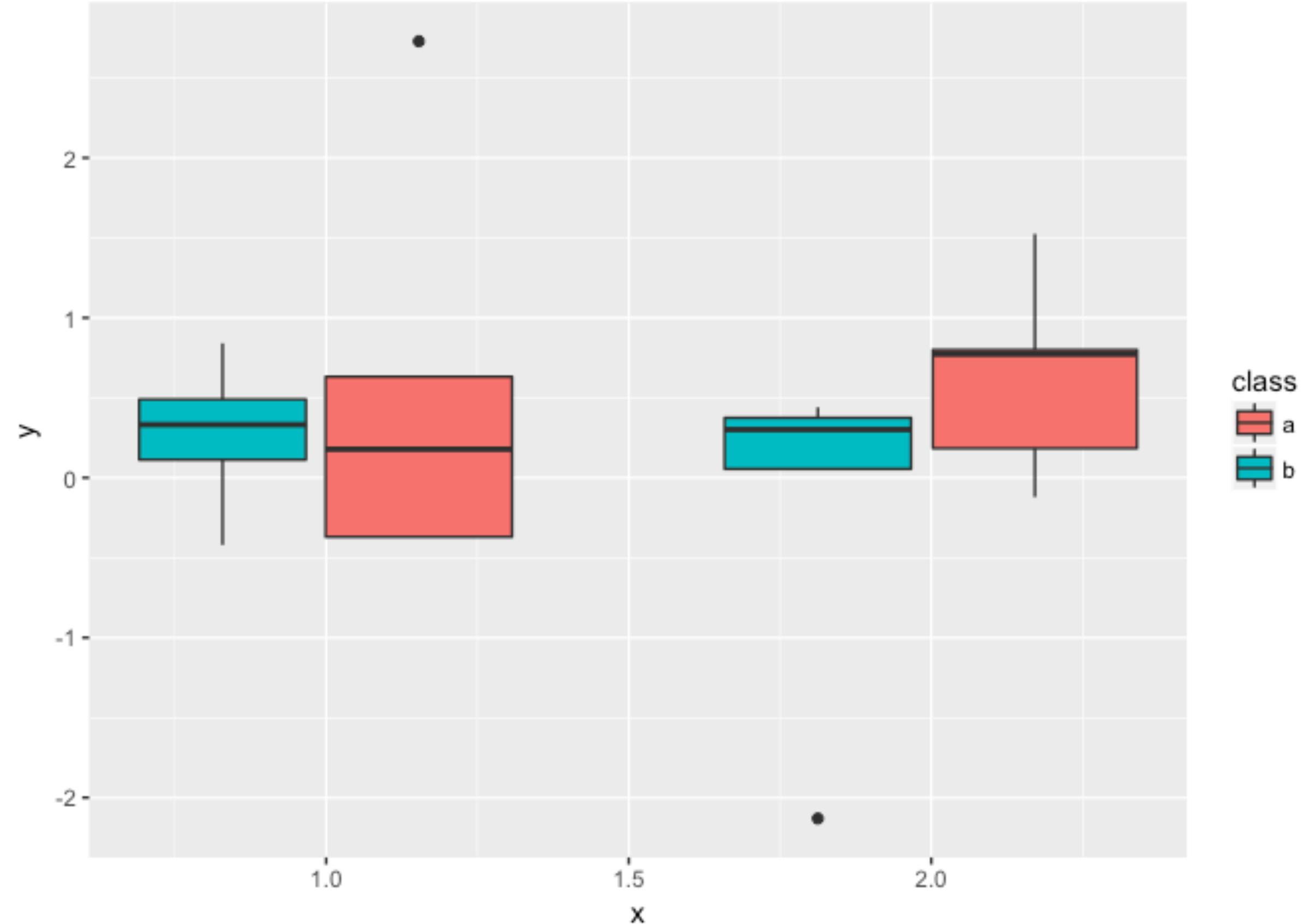
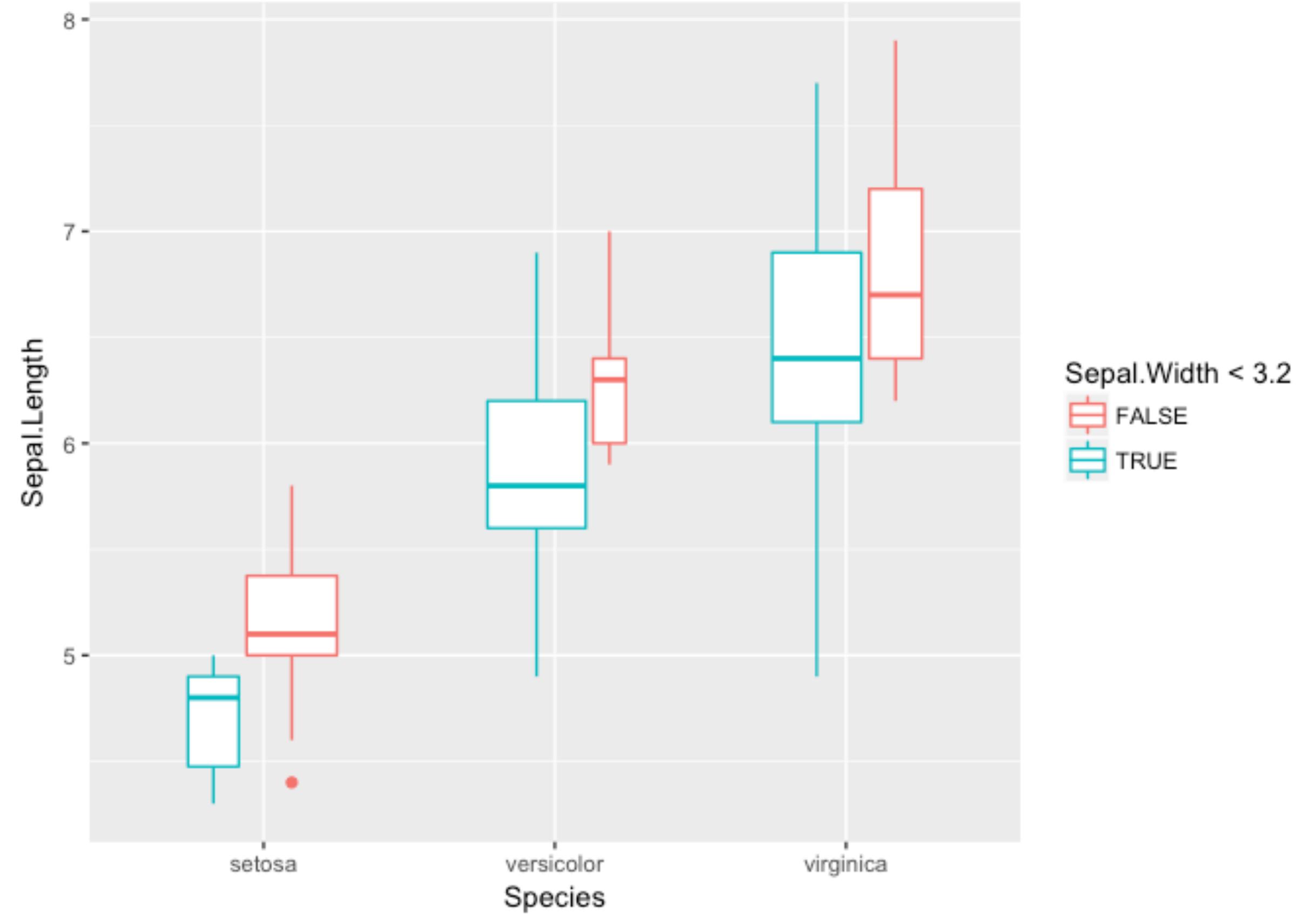
What's wrong with
this picture?

- Boxes in the wrong order
- Doesn't work for continuous x axes
- Incorrect scaling



HOW DO
YOU FIX IT?

Test many scenarios





HOW DO
YOU FIX IT?

Make it general

Can we extend to bars?

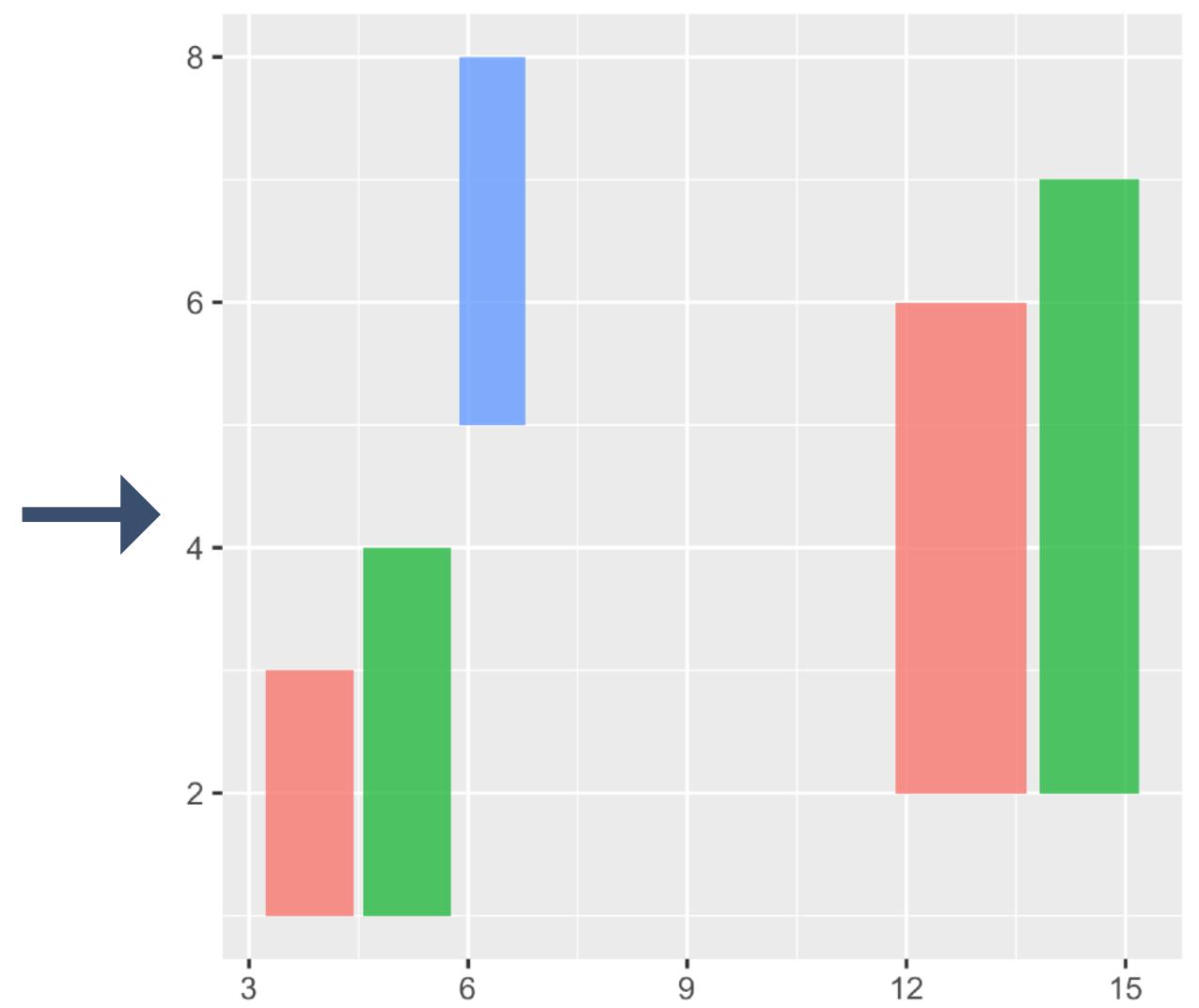
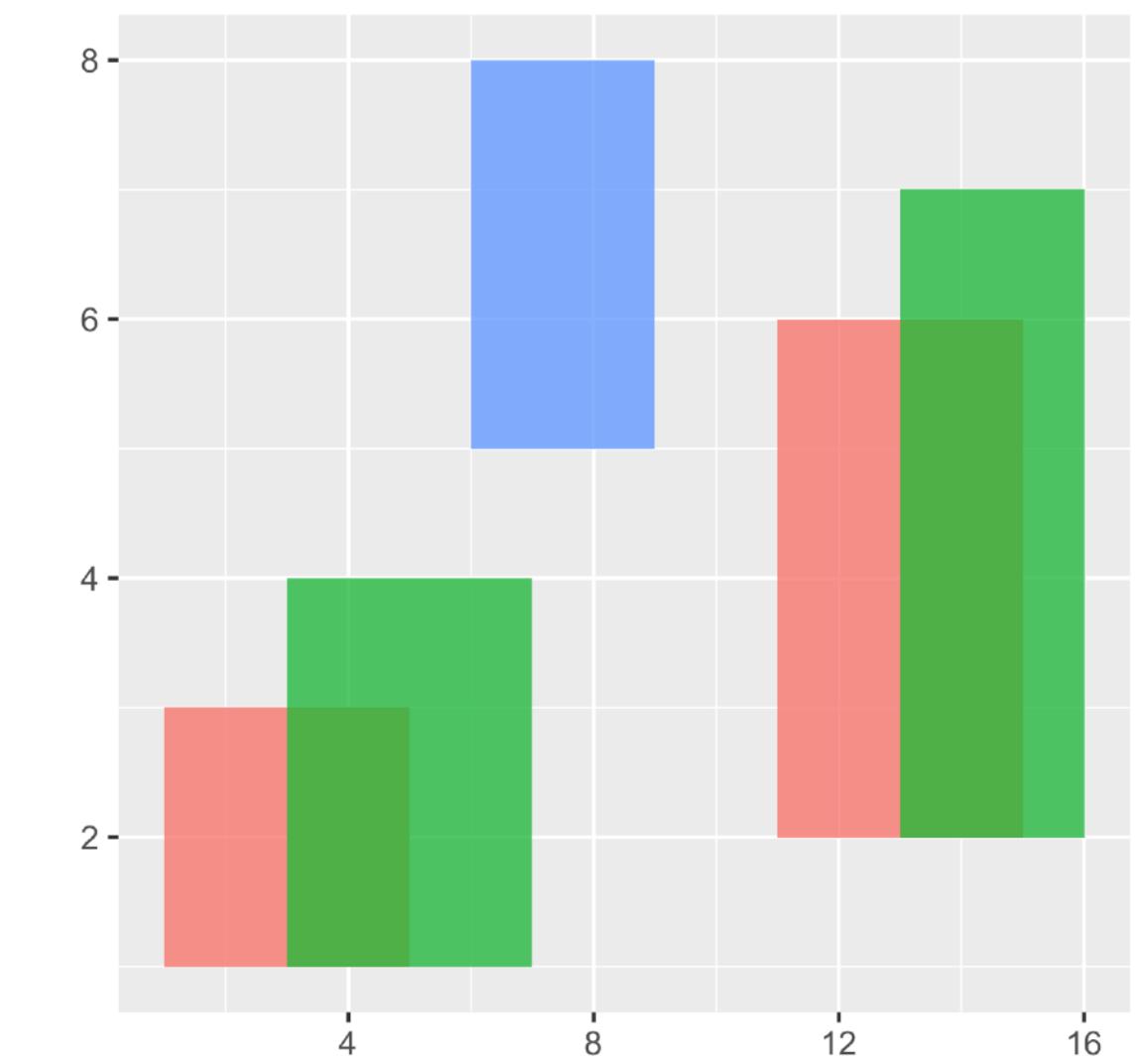
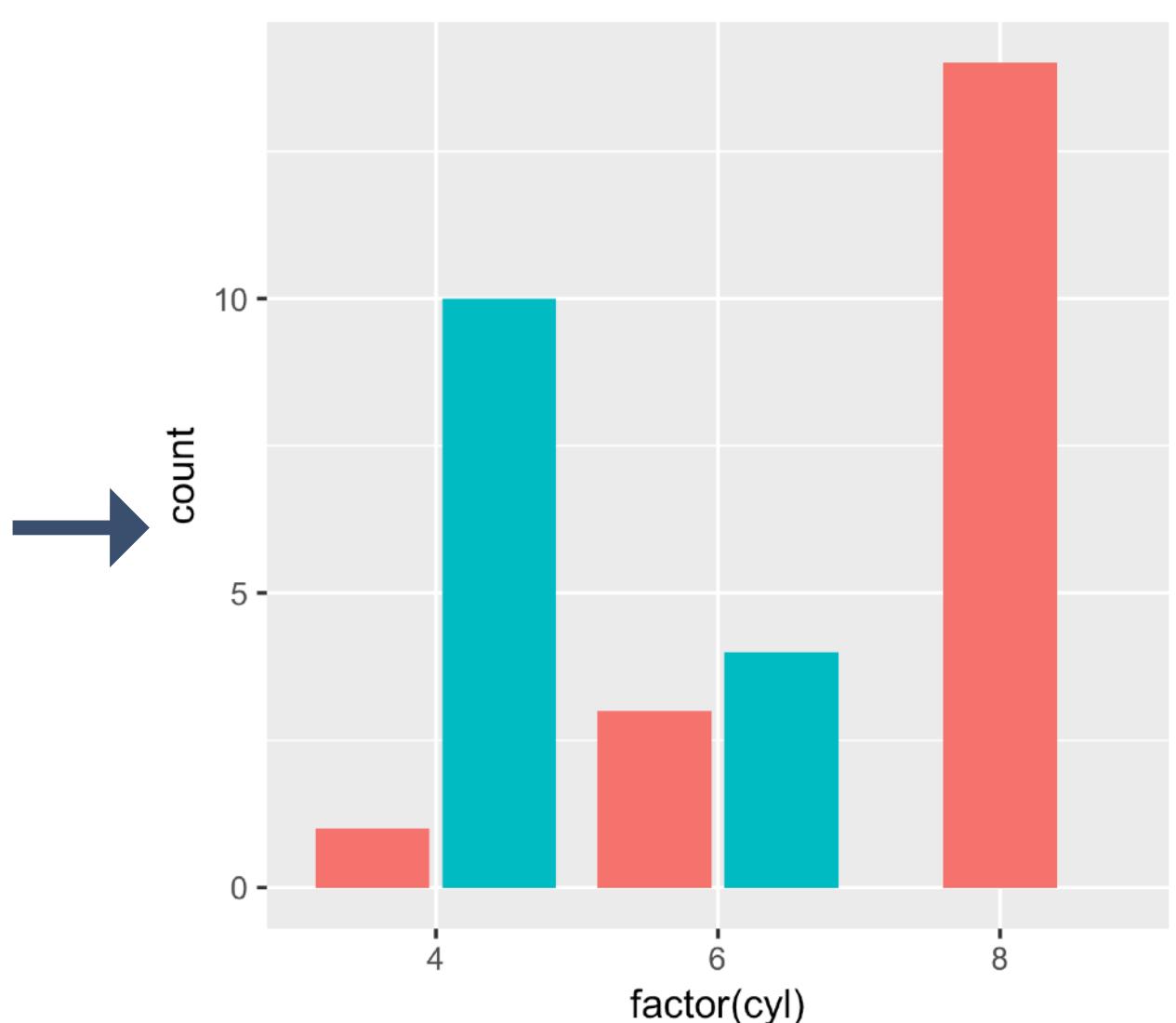
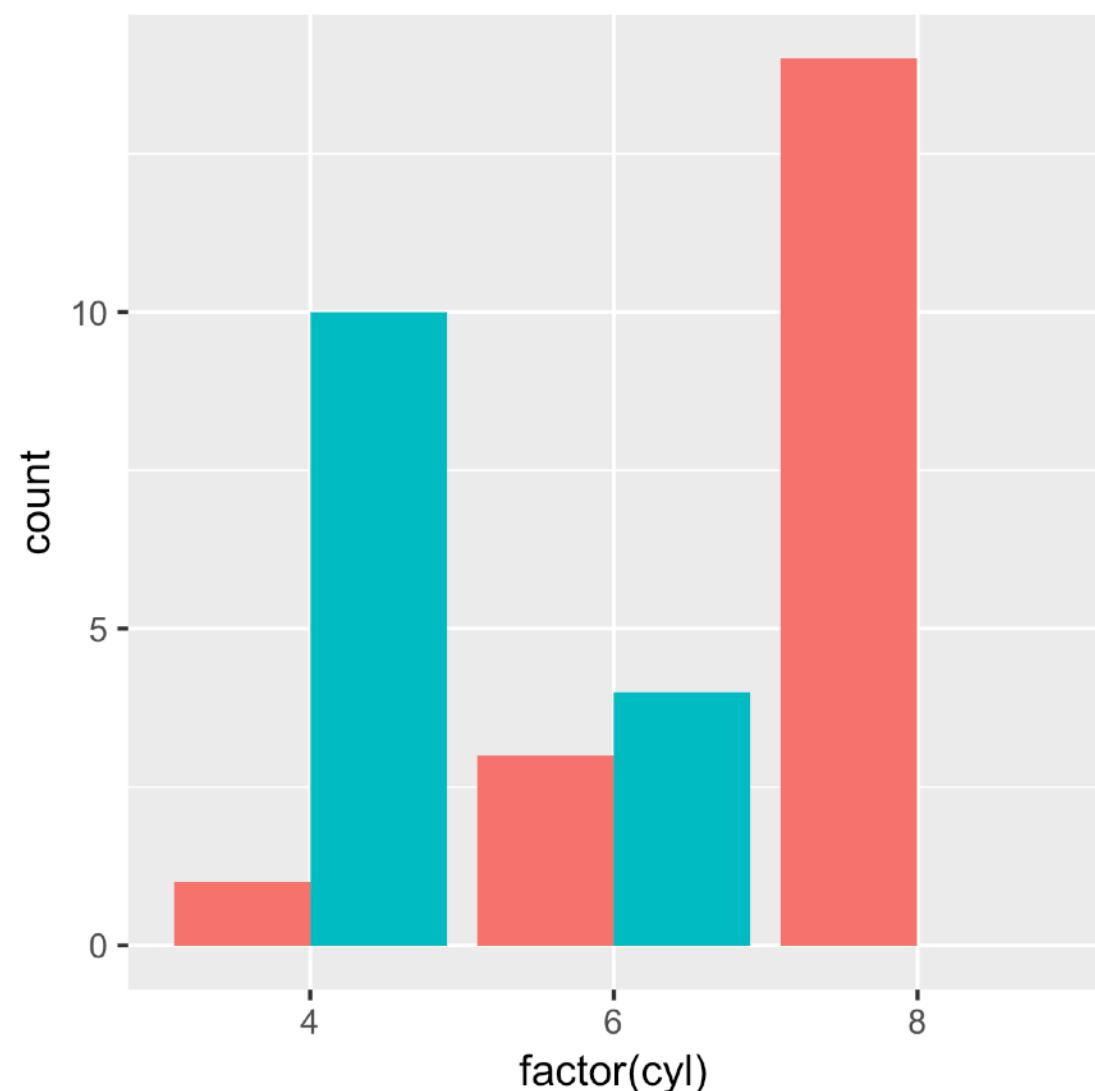
Arbitrary rectangles?

Can we solve other open issues?

POSITION_DODGE2()

Used for boxes, bars, rectangles

Finds overlap by comparing
xmin to previous xmax



HOW DO YOU
KNOW WHEN
YOU'RE DONE?

Again, test



Photo: Jia Ye

HOW DO YOU KNOW WHEN YOU'RE DONE?

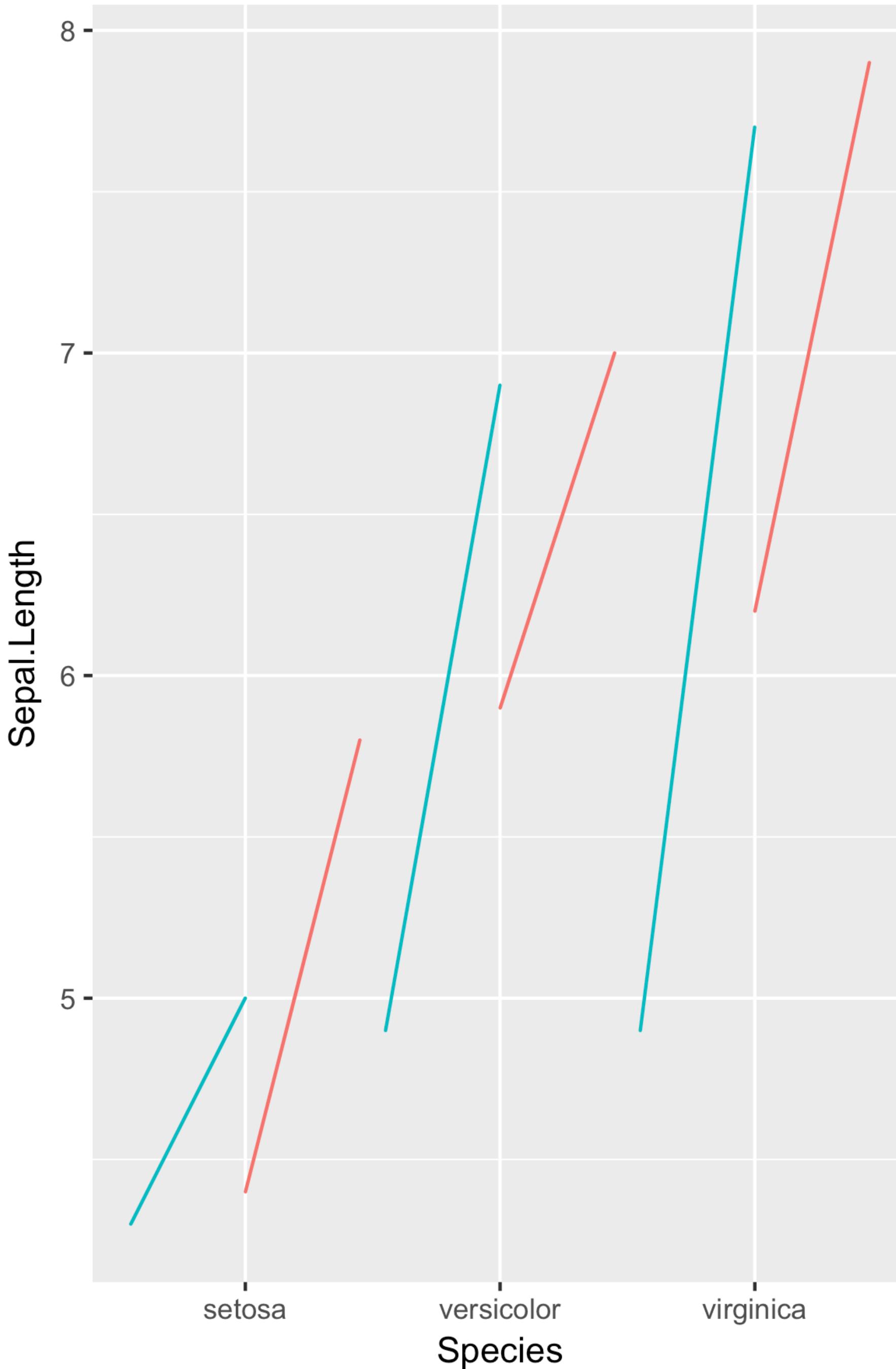
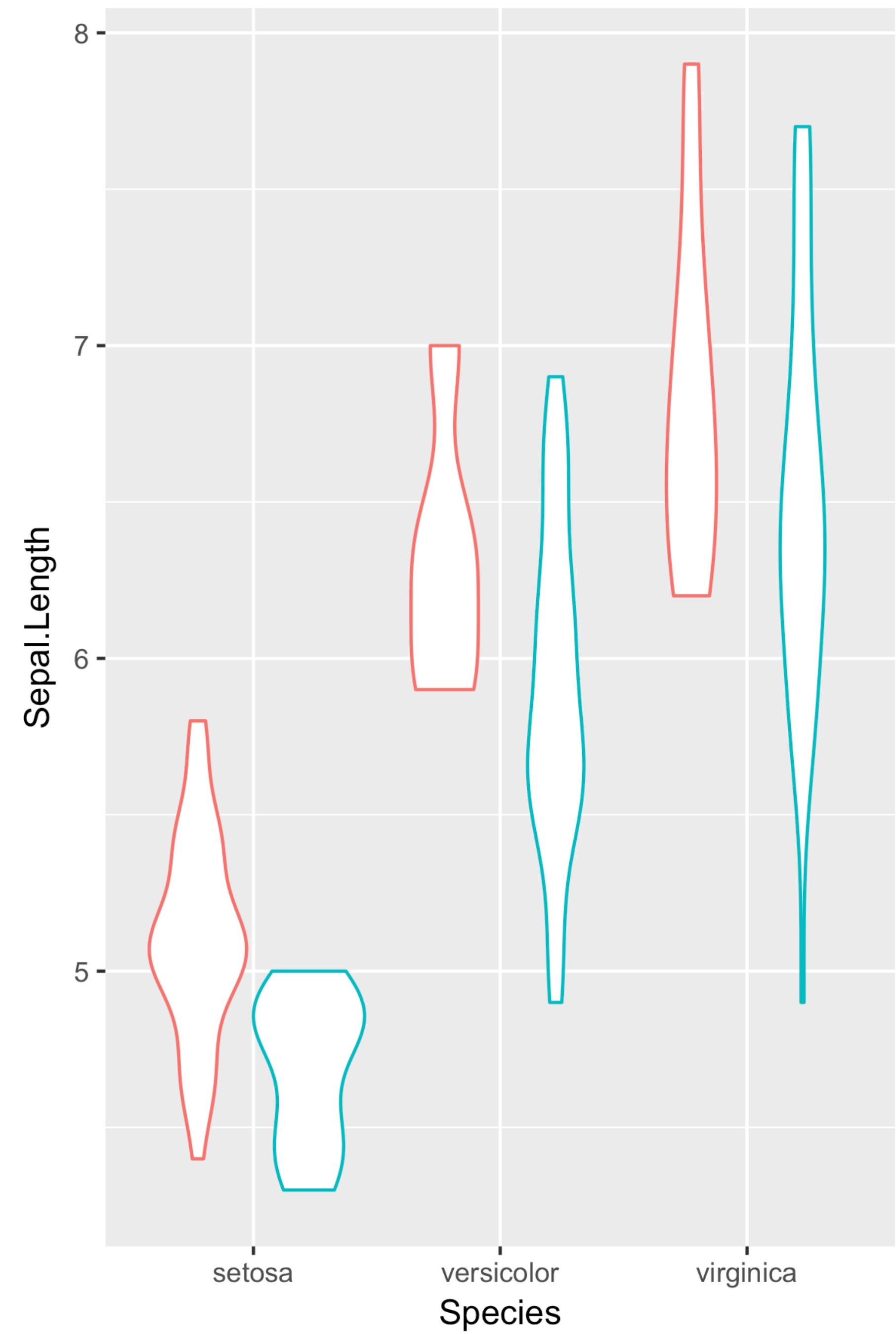
Don't let
perfect be the
enemy of good

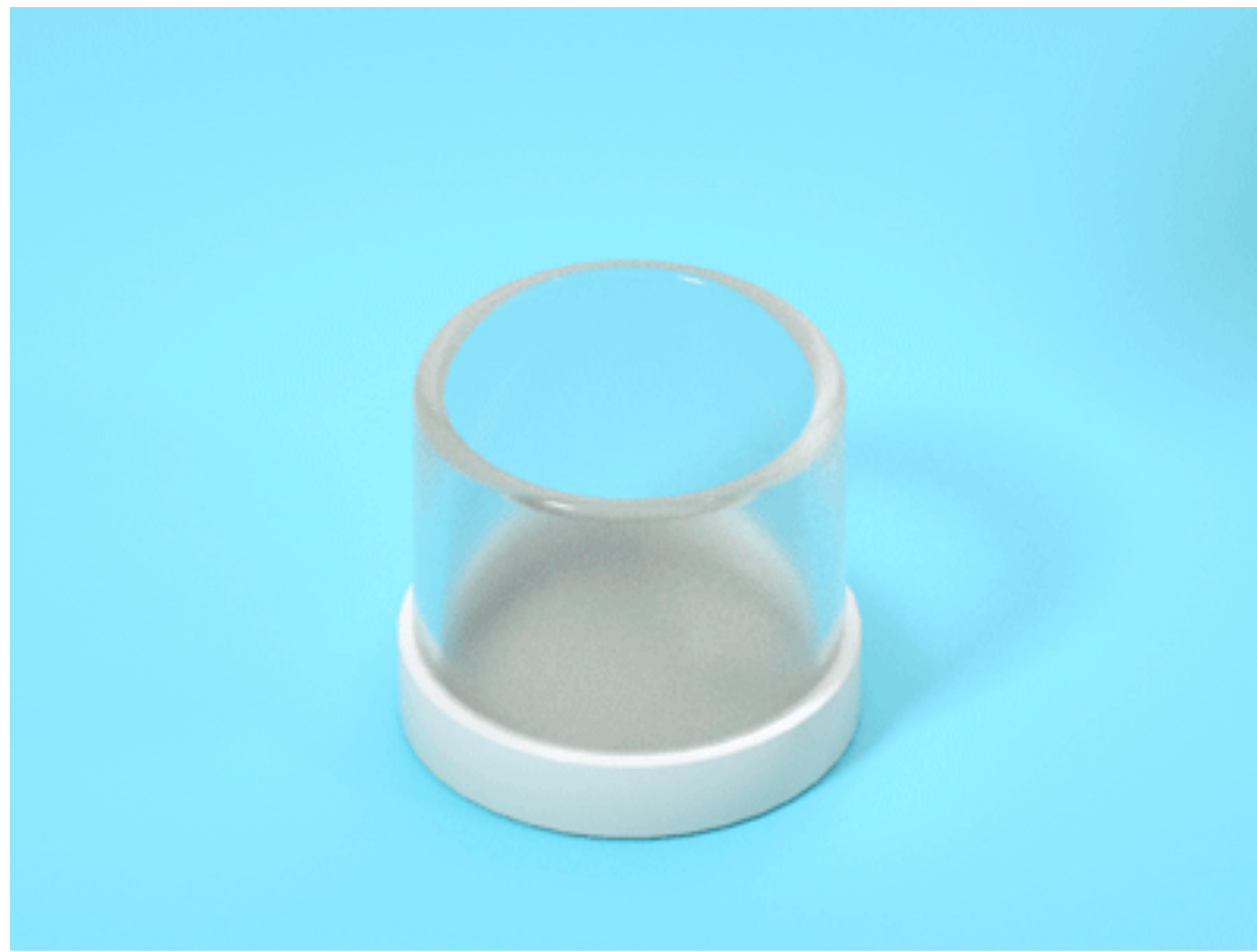


Photo: Jia Ye



CAN
POSITION_DODGE2()
REPLACE
POSITION_DODGE()?





<https://github.com/tidyverse/ggplot2/pull/2196>

- Isolate the problem
- Follow the trails
- Experiment
- Test many scenarios
- Make it general
- Don't let perfect be the enemy of good

Thank you

Thanks to @mcol for reporting this bug,
Hadley Wickham for repeated code
reviews, and Sean Kross and Karthik Ram
for feedback on this presentation.