Visualising a difference in mean between two groups isn’t as straightforward as it should. After all, it’s probably the most common quantitative analysis in science. There are two obvious options: we can either plot the data from the two groups separately, or we can show the estimate of the difference with an interval around it.

A swarm of dots is good because it shows the data, but it obscures the difference, and has no easy way to show the uncertainty in the difference. And, unfortunately, the uncertainty of the means within groups is not the same thing as the uncertainty of the difference between means.

An interval around the difference is good because it makes the plausible range of the difference very clear, but it obscures the range and distribution of the data.

Let’s simulate some fake data and look at these plots:

library(broom)

library(egg)

library(ggplot2)

data <- data.frame(group = rep(0:1, 20))

data$response <- 4 + data$group \* 2 + rnorm(20)

We start by making two clouds of dots. Then we estimate the difference with a simple linear model, and plot the difference surrounded by an approximate confidence interval. We can plot them separately or the *egg* package to put them together in two neat panels:

plot\_points <- ggplot() +

geom\_jitter(aes(x = factor(group), y = response),

data = data,

width = 0.1) +

xlab("Group") +

ylab("Response") +

theme\_bw()

model <- lm(response ~ factor(group), data = data)

result <- tidy(model)

plot\_difference <- ggplot() +

geom\_pointrange(aes(x = term, y = estimate,

ymin = estimate - 2 \* std.error,

ymax = estimate + 2 \* std.error),

data = result) +

ylim(-5, 5) +

ylab("Value") +

xlab("Coefficient") +

coord\_flip() +

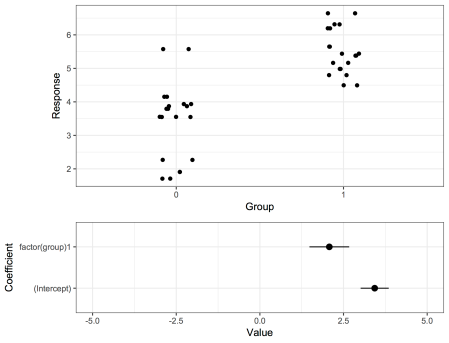
theme\_bw()

plot\_combined <- ggarrange(plot\_points,

plot\_difference,

heights = c(2, 1))

Here it is:

[](https://martinsbioblogg.files.wordpress.com/2019/01/blogpostplot.png)

But I had another idea. I am not sure whether it’s a good idea or not, but here it is: We put in the dots, and then we put in two lines that represent the smallest and the greatest difference from the approximate confidence interval:

offset <- (2 \* result$estimate[1] + result$estimate[2])/2

shortest <- result$estimate[2] - 2 \* result$std.error[2]

longest <- result$estimate[2] + 2 \* result$std.error[2]

plot\_both <- plot\_points +

geom\_linerange(aes(ymin = offset - shortest/2,

ymax= offset + shortest/2,

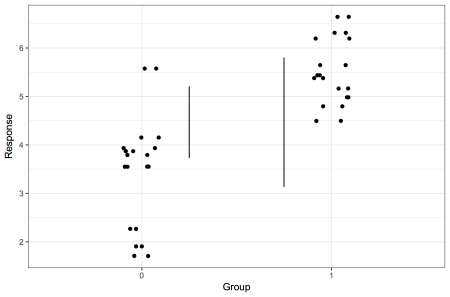
x = 1.25)) +

geom\_linerange(aes(ymin = offset - longest/2,

ymax= offset + longest/2,

x = 1.75)) +

theme\_bw()

[](https://martinsbioblogg.files.wordpress.com/2019/01/blogpostplot2.png)

I think it looks pretty good, but it’s not self-explanatory, and I’m not sure whether it is misleading in any way.