# Box Plot v.s. Violin Plot

[**HOME**](http://yufree.github.io/blog)[**APP**](http://yufree.github.com/blog/app/)[**ABOUT**](http://yufree.github.com/blog/about/)[**VITAE**](http://yufree.github.com/blog/vitae/)[**GUESTBOOK**](http://yufree.github.com/blog/guestbook/)[**SUBSCRIBE**](http://yufree.github.com/blog/feed/)

In a seminar I introduced the violin plot and showed the following figure(this example comes from the help document):

library(vioplot)

library(sm)

par(mfrow = c(1, 2))

mu <- 2

si <- 0.6

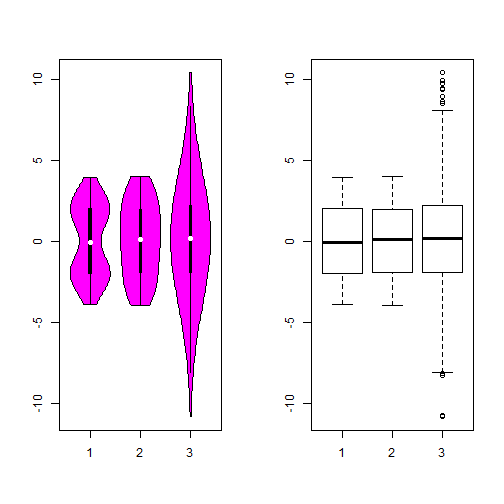
bimodal <- c(rnorm(1000, -mu, si), rnorm(1000, mu, si))

uniform <- runif(1000, -4, 4)

normal <- rnorm(2000, 0, 3)

vioplot(bimodal, uniform, normal)

boxplot(bimodal, uniform, normal)



So obviously, the violin plot can show more information than box plot. When we perform an exploratory analysis, nothing about the samples could be known. So the distribution of the samples can not be assumed to a normal distribution and usually when you get a big data, the normal distribution will show some out liars in box plot. Referring to the paper by Hintze, J. L. and R. D. Nelson (1998), the violin plot combines the box plot and the density trace, so it seems that the box plot may give the place to the violin plot and I said this in the seminar from a viewpoint of environmental science. But after the seminar, I really doubt that no environmental scientists use this plot. Of course, the violin plot is young comparing with the box plot introduced by Tukey(1977), but there also exist some reasons which stop the spread of violin plot. Here I list it as follows:

* the violin plot can’t show a better curve with small samples. In Hintze’s paper, he thought a smooth curve with at least 30 observations. But the box plot may stand for a smaller observations. Also the bandwidth need to be chosed carefully.
* the modification box plot could show the number of observations in the groups using the var width while the violin plot couldn’t. When we make some comparison between different groups, the violin plot will hide this information.
* Another problem is the notch in the box plot to compare the median. In the violin plot, we get a better understanding of distribution of violin plot but less with comparisone with ‘strong evidence’(Chambers et al., 1983, p. 62).

Those were nitpick reasons but I think if someone just want to show the violin plot instead of box plot, he need to know the details. Nowadays, it is easy to use new concepts to confusing the readers, we need more thoughts about the nature. Here is a example: there are numbers people who thought the box plot show the mean…