

ToothGrowth 과제 풀이

YL

5-28-2019

Crampton은 음식물 내 아스코르브산이 용량 별로 실험동물의 치아 성장에 미치는 효과를 알고자 총 60마리의 기니피그를 골라 비타민캡슐과 오렌지주스로 나누어 실험하였다^a R의 내장 자료철 **ToothGrowth**가 이 실험자료를 가지고 있으며 다음처럼 변수 **len** (numeric. 치아 길이), **supp** (factor. 오렌지주스는 OJ, 비타민캡슐은 VC로 코딩), **dose** (numeric. 용량)로 구성되어 있다:

```
str(ToothGrowth)

## 'data.frame': 60 obs. of 3 variables:
## $ len : num 4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
## $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
## $ dose: num 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

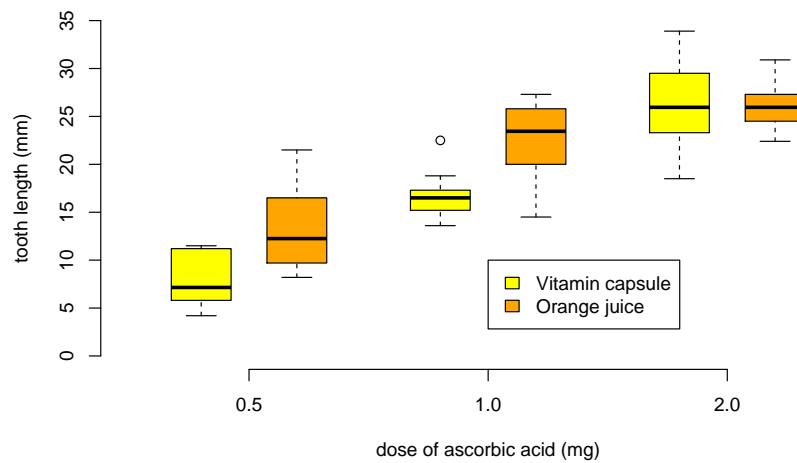
^aCrampton, E. W. (1947). The growth of the odontoblast of the incisor teeth as a criterion of vitamin C intake of the guinea pig. The Journal of Nutrition, 33(5), 491–504. doi: 10.1093/jn/33.5.491.

이 **ToothGrowth** 자료에서 치아의 성장을 아스코르브산 용량에 따라 도해하면서 오렌지주스와 비타민캡슐을 색깔로 대별할 수 있도록 하되, R 기본함수인 **boxplot()**을 이용하라.

정답 0

```
boxplot(len ~ dose, data = ToothGrowth,
        subset = supp == "VC", col = "yellow",
        boxwex = 0.25, at = 1:3 - 0.2,
        xlab = "dose of ascorbic acid (mg)",
        ylab = "tooth length (mm)",
        xlim = c(0.5, 3.5), ylim = c(0, 35),
        axes = FALSE) # yaxs = "i"
boxplot(len ~ dose, data = ToothGrowth, add = TRUE,
        subset = supp == "OJ", col = "orange",
        boxwex = 0.25, at = 1:3 + 0.2,
        axes = FALSE)
axis(1, at=1:3, labels = c("0.5", "1.0", "2.0"))
axis(2)
```

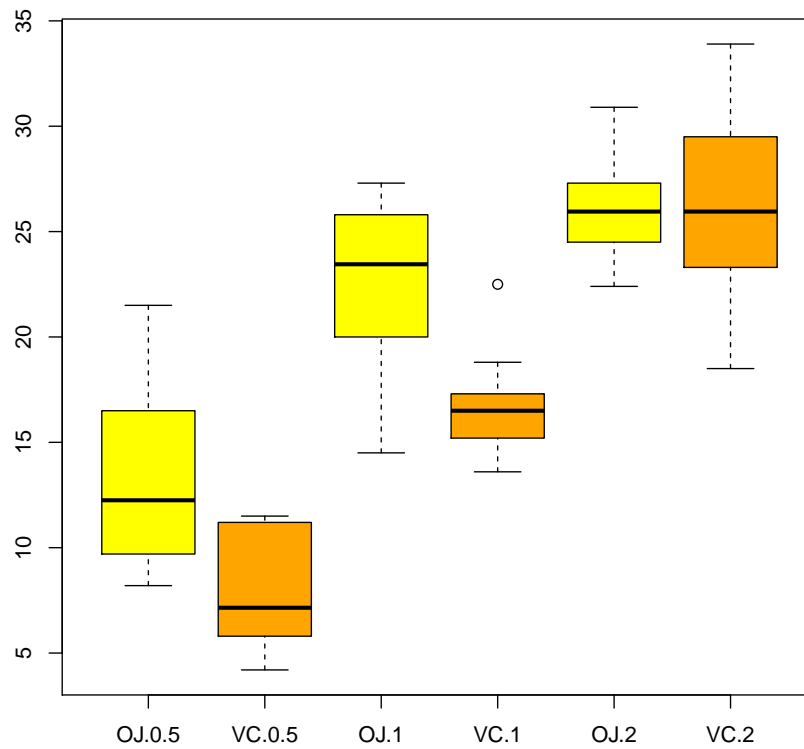
```
legend(2, 10, c("Vitamin capsule", "Orange juice"),
      fill = c("yellow", "orange"))
```



해법 1

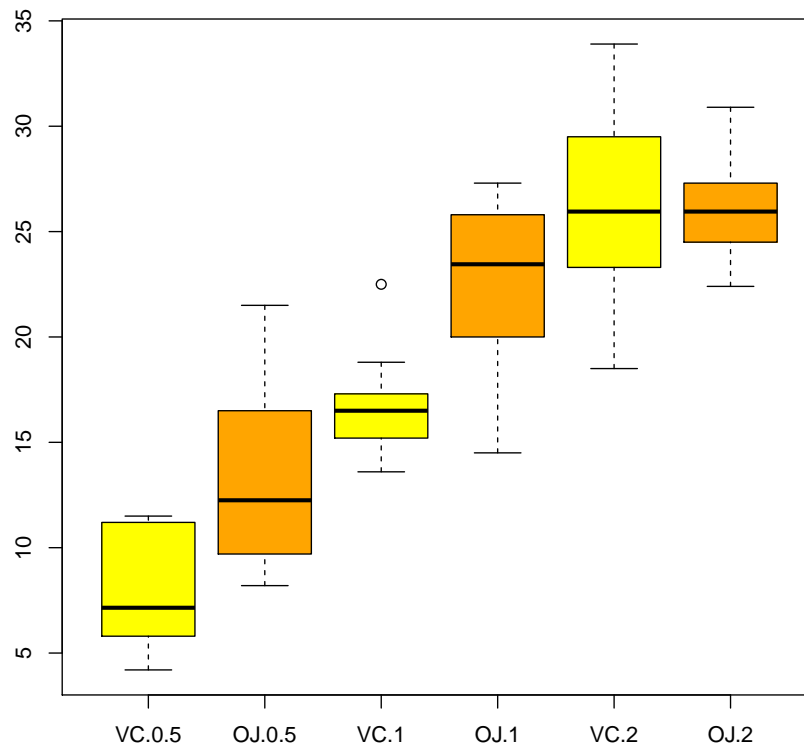
자료철에서 `supp` 변수의 레벨이 알파벳 순으로 되어 있으므로 조정하지 않고 그리면 OJ가 VC를 앞서게 하는 실수를 범한다.

```
# solution 1
TG <- ToothGrowth
boxplot(len ~ supp:dose, data = TG, col = rep(c("yellow", "orange"), 3)) #
```



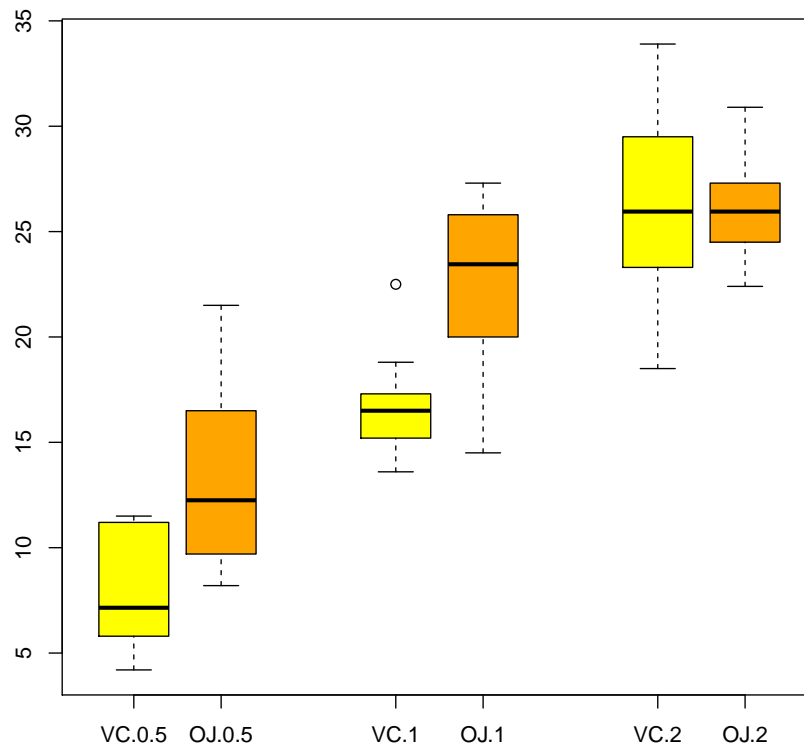
다음처럼 supp의 레벨을 VC < OJ로 바꾼 자료철 TG를 만들어서 TG에서 작업한다.

```
# 1.1
TG$supp <- factor(TG$supp, levels = c("VC", "OJ"))
boxplot(len ~ supp:dose, data = TG, col = rep(c("yellow", "orange"), 3)) #
```



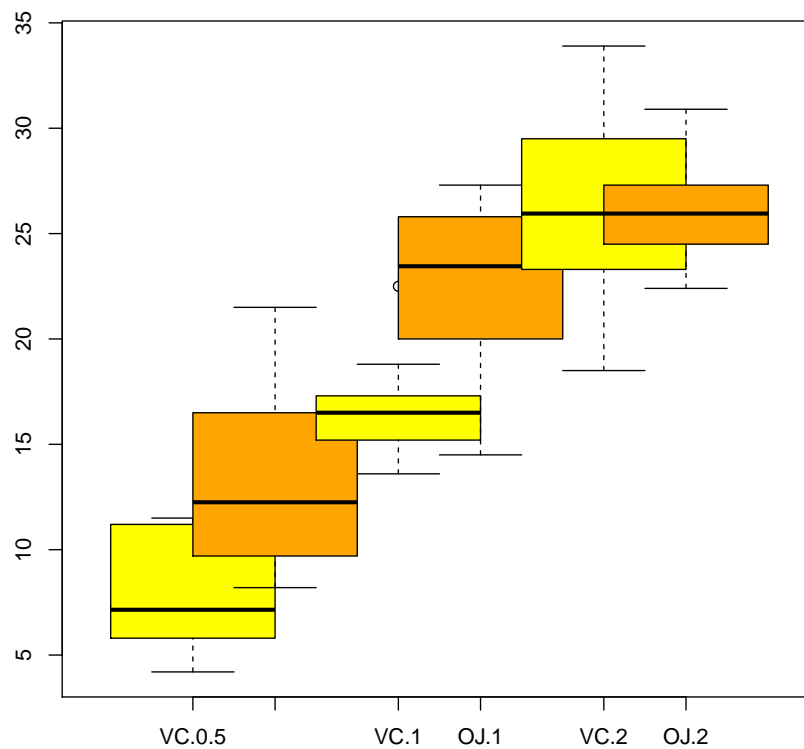
원하는 대로 순서는 정렬되었지만 박스 사이의 거리는 여전히 마음에 들지 않는다. 위치를 강제 지정하기 위해서 `boxplot()` 내 인수 `at`을 써 보았다.

```
boxplot(len ~ supp:dose, data = TG,
        col = rep(c("yellow", "orange"), 3),
        at = c(1, 2, 4, 5, 7, 8)) #
```



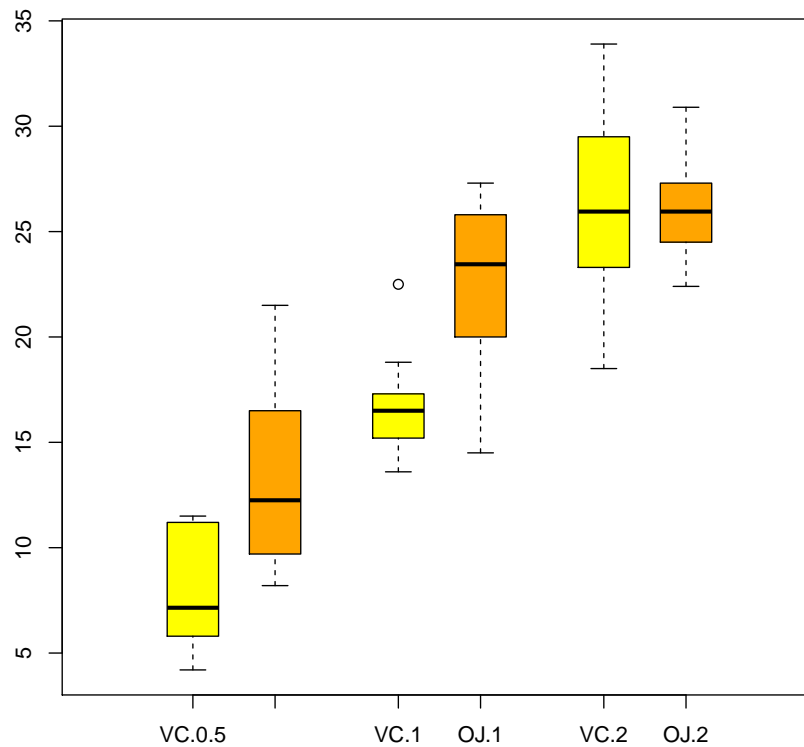
dose 사이의 간격이 너무 멀어서 at을 새로 지정한다. 1, 2, 3을 중심으로 0.2를 가감한 꼴이다.

```
position <- c(0.8, 1.2, 1.8, 2.2, 2.8, 3.2)
boxplot(len ~ supp:dose, data = TG,
        col = rep(c("yellow", "orange"), 3), at = position) #
```



이렇게 했더니 box의 간격 때문에 겹쳐서 식별이 되지 않는다 boxwex 인수를 조정해야 한다.

```
boxplot(len ~ supp:dose, data = TG,
  col = rep(c("yellow", "orange"), 3),
  at = position, boxwex = 0.25) #
```

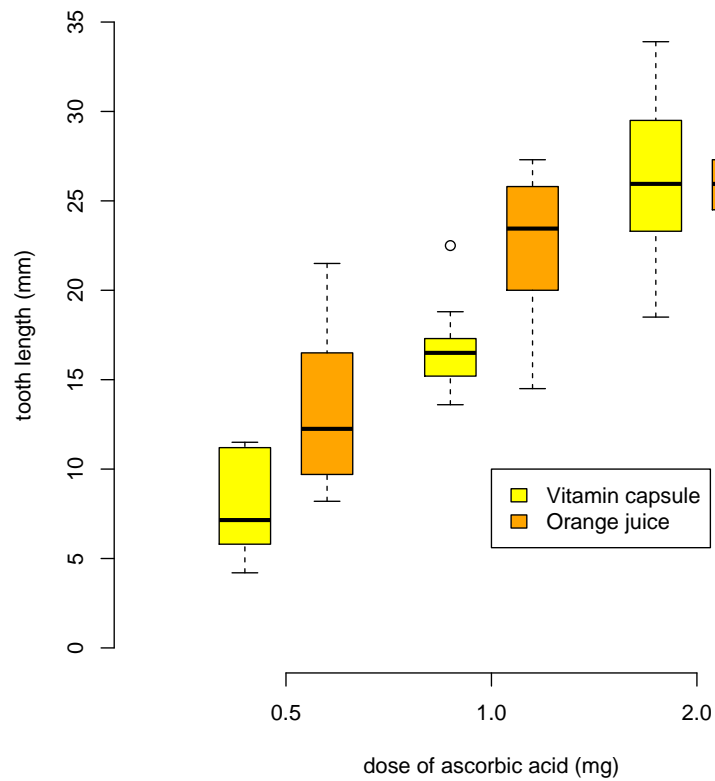


축 이름과 축 간격 이름을 조정하기 위해서 디폴트 기능을 쓰지 않고 `axis()` 를 재조정했다.

```
boxplot(len ~ supp:dose, data = TG,
        col = rep(c("yellow", "orange"), 3),
        at = position, boxwex = 0.25,

        # tidying ~
        ylab = "tooth length (mm)",
        xlab = "dose of ascorbic acid (mg)",
        ylim = c(0, 35),
        axes = FALSE) #

axis(2)
axis(1, at = 1:3, labels = c("0.5", "1.0", "2.0"))
legend(2, 10, c("Vitamin capsule", "Orange juice"),
      fill = c("yellow", "orange"))
```



해법 2

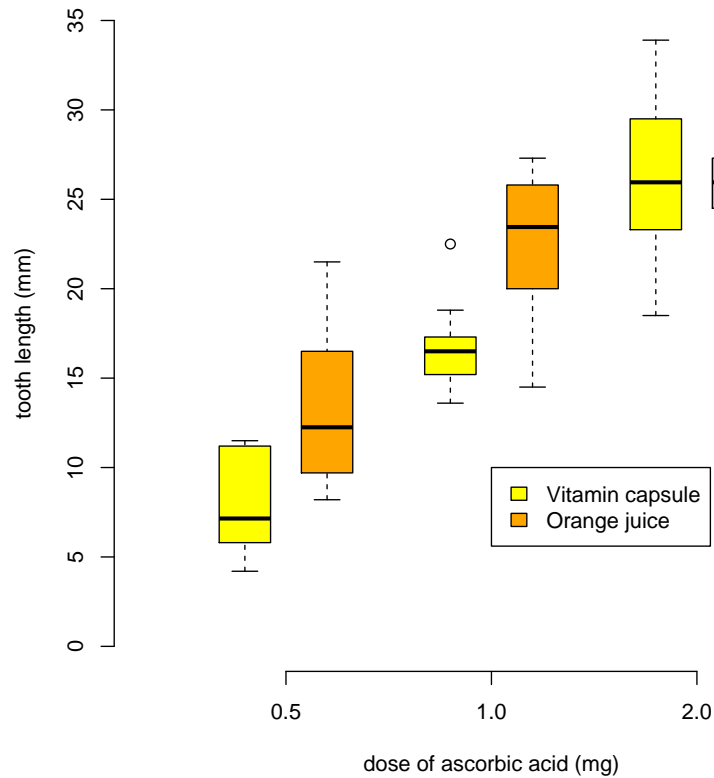
boxplot의 at 인수를 잘 이해하면 supp 변수의 레벨을 바꾸지 않고도 원하는 위치에 그래프를 그릴 수 있다. 즉 위치가 0.8, 1.2, 1.8, ... 식으로 순차적인 진행을 하는 것이 아니라 1.2, 0.8, 2.2, ... 식으로 엇갈려서 찍는 것이다.

```
position <- c(1.2, 0.8, 2.2, 1.8, 3.2, 2.8)
boxplot(len ~ supp:dose, data = ToothGrowth,
        col = rep(c("orange", "yellow"), 3),
        at = position,
        boxwex = 0.25,
        ylab = "tooth length (mm)",
        xlab = "dose of ascorbic acid (mg)",
        ylim = c(0, 35),
        axes = FALSE) #

axis(2)
axis(1, at = 1:3, labels = c("0.5", "1.0", "2.0"))
legend(2, 10, c("Vitamin capsule", "Orange juice"),
```



```
fill = c("yellow", "orange"))
```



결과물은 해법 1과 동일하다.

해법 3

특별한 해법이라고 부르기엔 부족하지만 두 개의 `boxplot` 호출로 `vc`와 `oj`를 완전히 따로 그리는 방법이 있다. 자료철을 분철(subsetting)하는 방식이다. 이렇게 하면 그릴 위치를 별개로 지정할 수 있어서 생각이 복잡해지지 않는다는 장점이 있다.

```
boxplot(len ~ dose, data = ToothGrowth,
        subset = (supp == "VC"),
        col = "yellow", at = 0.8:3, boxwex = 0.25,
        ylim = c(0, 35),
        ylab = "tooth length (mm)",
        xlab = "dose of ascorbic acid (mg)",
        axes = FALSE) # !!

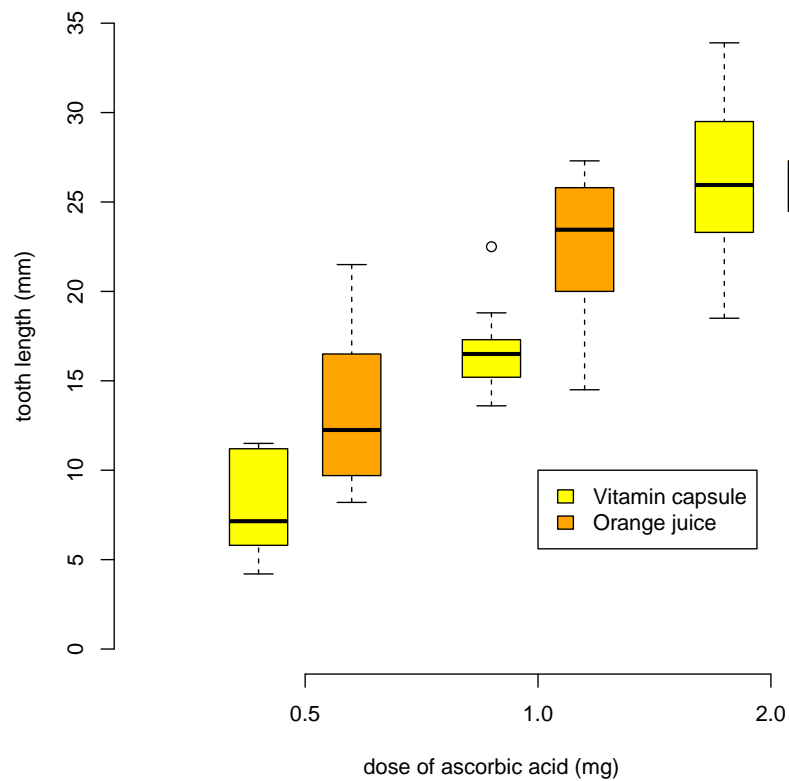
boxplot(len ~ dose, data = ToothGrowth,
        subset = (supp == "OJ"),
```

```

col = "orange", at = 1.2:4, boxwex = 0.25,
axes = FALSE,
add = TRUE) # !!

axis(2)
axis(1, at = 1:3, labels = c("0.5", "1.0", "2.0"))
legend(2, 10, c("Vitamin capsule", "Orange juice"),
fill = c("yellow", "orange"))

```



이렇게 그릴 때 주의할 점은, 두 boxplot 모두의 axes = FALSE로 놓아야 한다는 점이다.

해법 4

변수 `supp`의 레벨에 빈 레벨 하나를 삽입함으로써 빈 공간을 확보하는 방법도 있다. 이렇게 새 자료철 `TG`를 만든다. 아무 내용도 없는 레벨 `NN`이 확보되었다.

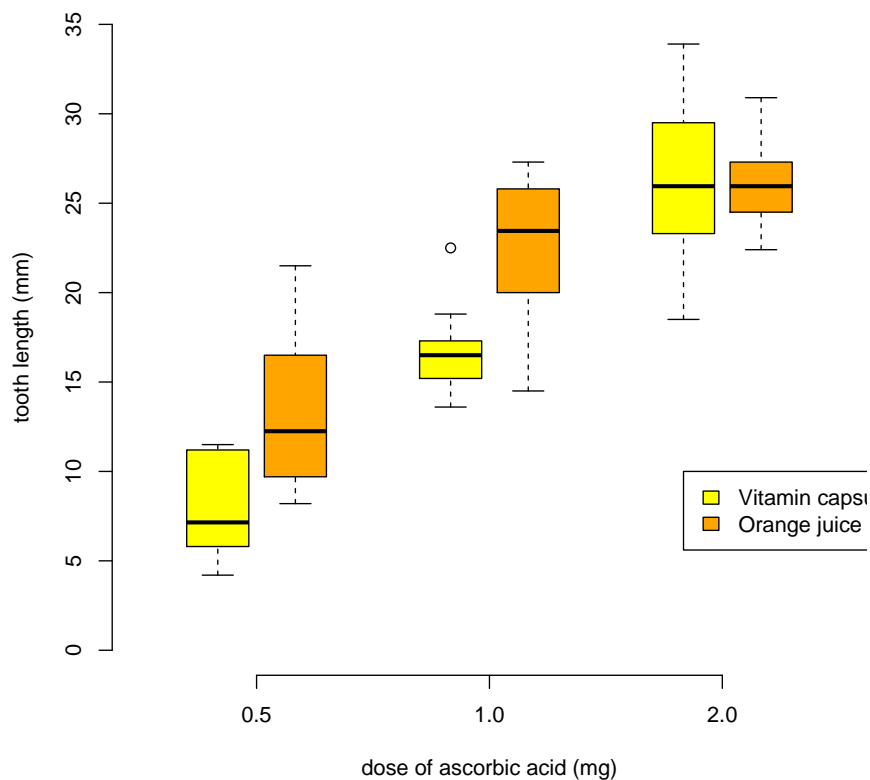
```

TG <- ToothGrowth
TG$supp <- factor(TG$supp, levels = c("VC", "OJ", "NN"))

```

여기부터 그리는 과정은 순조로운데 이렇게 그려진 그래프의 가로축의 틱(tick) 이름을 쓸 위치는 고민 끝에 결정해야 한다. 즉, 한 dose 내 세 개의(두 개와 빈 것 하나) 그래프가 1, 2, 3에 그려지므로 3은 비어 있으므로 첫 tick의 위치는 1.5가 된다. 같은 방식으로 두 번째는 4.5, 세 번째는 7.5이다. 코드 내에 주석 here!!이다.

```
boxplot(len ~ supp:dose, data = TG,
        col = rep(c("yellow", "orange", "white"), 3),
        ylim = c(0, 35),
        xlim = c(0, 9),
        ylab = "tooth length (mm)",
        xlab = "dose of ascorbic acid (mg)",
        axes = FALSE)
axis(2)
# Please note 'at' positions
axis(1,
     at = c(1.5, 4.5, 7.5), # here!!
     labels = c("0.5", "1.0", "2.0"))
legend(7, 10, c("Vitamin capsul", "Orange juice"),
      fill = c("yellow", "orange"))
```



조별 사정

조A

에러 발생. 두 곳의 에러 발생 부분을 고쳐서 실행하였지만 supp 레벨을 정리하지 못했고 박스의 모으기를 실패한 그래프가 나옴.

```
boxplot(len~supp*dose,
        xlab=" dose of ascorbic acid(mg)",
        ylab=" tooth length(mm)"
        , data=ToothGrowth, col=c("orange","yellow"),
        axes=FALSE, )

## Error in boxplot.default(split(mf[[response]], mf[-response],
drop = drop, : argument is missing, with no default

axis(2)

## Error in axis(2): plot.new has not been called yet

axis(1, label=c(0.5,1.0,2.0), at=c(1.5,3.5,5.5))

## Error in axis(1, label = c(0.5, 1, 2), at = c(1.5, 3.5, 5.5)):
plot.new has not been called yet

legend(3.5,12,pch=0,col=c("orange","yellow"),
       c("orange juice","vitamin capsule"))

## Error in strwidth(legend, units = "user", cex = cex, font
= text.font): plot.new has not been called yet
```

조B

자료를 분할해서 그린 예로서 해법3과 동일한 접근법을 취했다. 다만, 가로축의 틱 이름이 두 박스의 중심부가 아닌 OJ 쪽으로 치우쳐서 안정감이 없다.

```
head(ToothGrowth, package = "nlme")

##      len supp dose
## 1  4.2    VC  0.5
## 2 11.5    VC  0.5
## 3  7.3    VC  0.5
## 4  5.8    VC  0.5
## 5  6.4    VC  0.5
## 6 10.0    VC  0.5

head(ToothGrowth, 5)

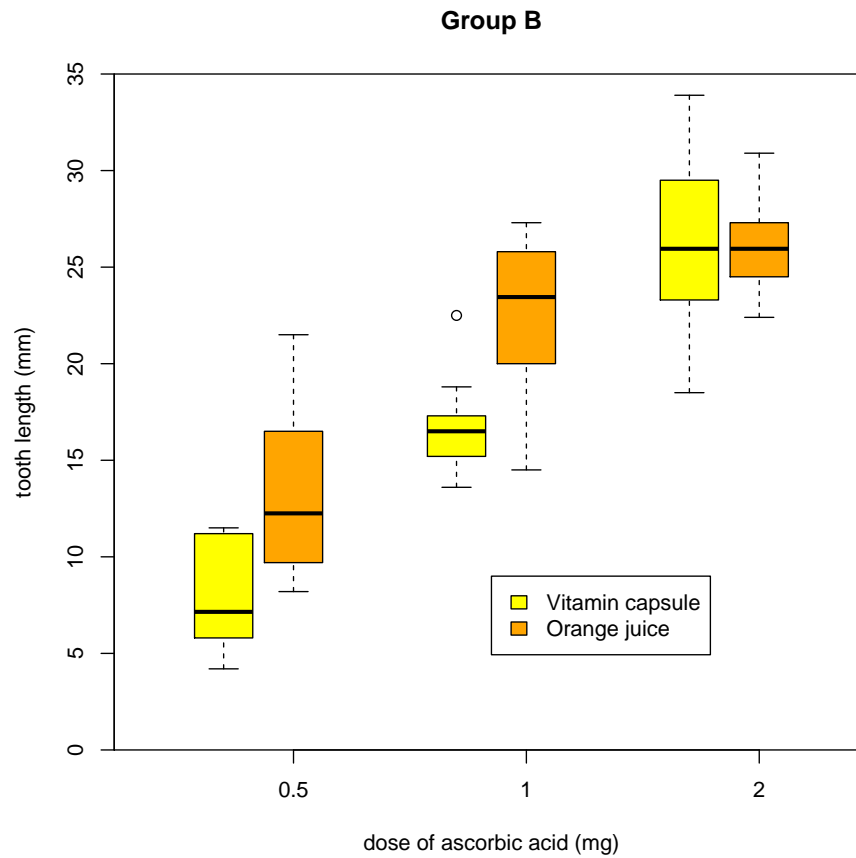
##      len supp dose
## 1  4.2    VC  0.5
## 2 11.5    VC  0.5
## 3  7.3    VC  0.5
```

```
## 4  5.8  VC  0.5
## 5  6.4  VC  0.5

str(ToothGrowth)

## 'data.frame': 60 obs. of  3 variables:
## $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
## $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
## $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...

boxplot(len ~ dose, data = ToothGrowth,
        main = "Group B",
        boxwex = 0.25, at = 1:3 - 0.15,
        subset = supp == "VC", col = "yellow",
        xlab = "dose of ascorbic acid (mg)",
        ylab = "tooth length (mm)",
        xlim = c(0.5, 3.5), ylim = c(0, 35), yaxs = "i", axes=FALSE)
boxplot(len ~ dose, data = ToothGrowth, add = TRUE,
        boxwex = 0.25, at = 1:3 + 0.15,
        subset = supp == "OJ", col = "orange")
legend(2, 9, c("Vitamin capsule", "Orange juice"),
      fill = c("yellow", "orange"))
```



조C

이 역시 자료철을 분리하여 VC와 OJ를 따로 그렸으나 가로축 틱 이름 정리에 실패했고 가로, 세로축 축이름의 단위가 빠졌다.

```
str(ToothGrowth)

## 'data.frame': 60 obs. of 3 variables:
## $ len : num 4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
## $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
## $ dose: num 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...

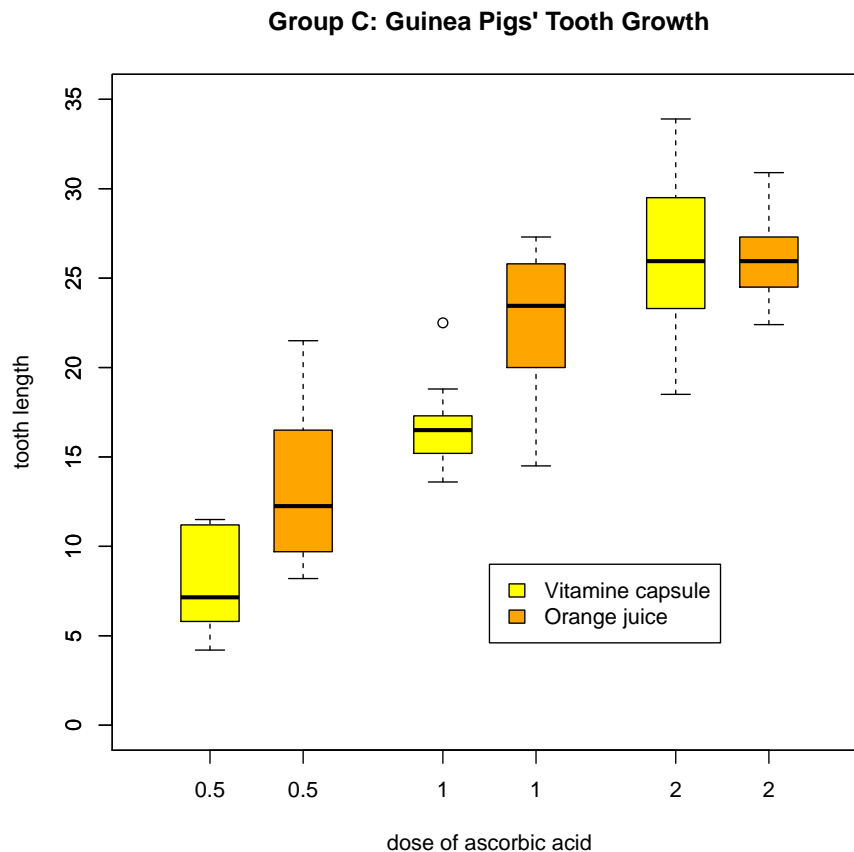
boxplot(len ~ dose, data = ToothGrowth,
        boxwex = 0.25, at = 1:3 - 0.2,
        col = "yellow", main = "Group C: Guinea Pigs' Tooth Growth",
        subset = supp == "VC",
        xlab = "dose of ascorbic acid",
        ylab = "tooth length",
        xlim = c(0.5, 3.5), ylim = c(0, 35))
```

```

boxplot(len ~ dose, data = ToothGrowth,
        add = TRUE,
        boxwex = 0.25, at = 1:3 + 0.2,
        subset = supp == "OJ", col = "orange")

legend(2, 9, c("Vitamine capsule", "Orange juice"),
       fill = c("yellow", "orange"))

```



조D

이 조는 `supp`의 요인 레벨을 정리하지 않고 그래프를 그렸으며 요일별 색상도 반대로 그렸다. 가로축 틱 위치는 제대로이나 `dose`간 군집화가 되지 않았다. 레전드에서 지시된 색상 안내가 박스의 안쪽이 아니라 경계선만으로 되었다.

```

str(ToothGrowth)

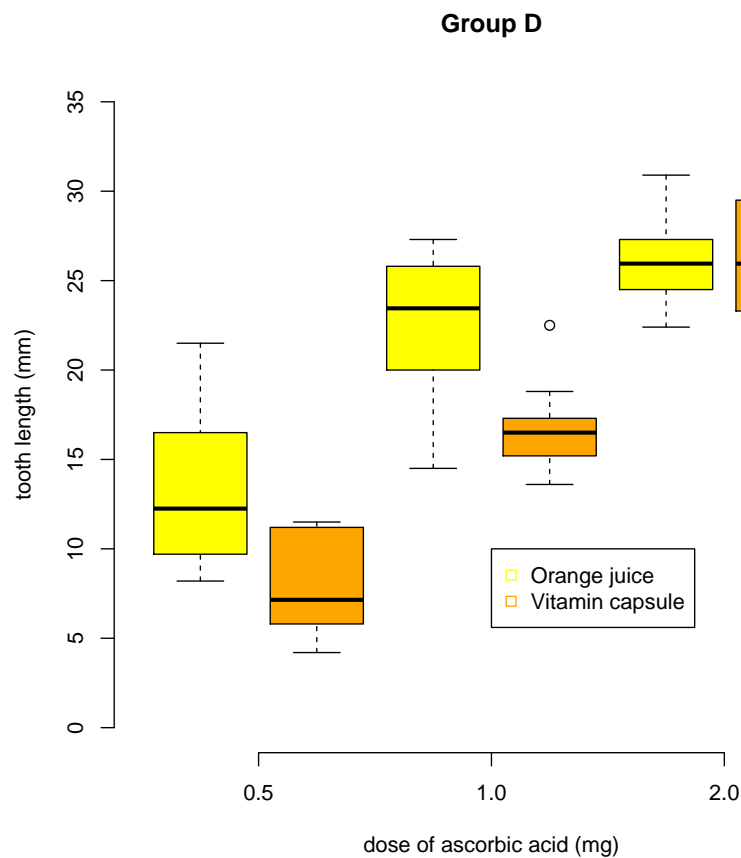
## 'data.frame': 60 obs. of 3 variables:
## $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
## $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
## $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...

```

```

lenAsMatrix <- matrix(ToothGrowth$len, byrow=TRUE, nrow =4)
boxplot( len~supp*dose,ylim=c(0,35), ToothGrowth,
  col =rep(c("yellow", "orange"),3),
  xlab = "dose of ascorbic acid (mg)",
  ylab = "tooth length (mm)", axes=FALSE,
  main = "Group D")
axis(2)
axis(1, label =c("0.5","1.0","2.0"), at= 1:3*2-0.5)
legend(3.5,10, col=c("yellow", "orange"),
  c("Orange juice","Vitamin capsule"), pch=0)

```



조E

박스의 폭이 다른 조에 비해서 넓고 dose 군집별로 묶음 처리를 하지 않았다. supp의 요인 레벨을 변환하는 데 rev() 함수를 사용하여 코드를 간결화한 점은 주목할 만하다.

```

boxplot(len ~ rev(supp) * dose, data = ToothGrowth,
  col = rep(c("yellow", "orange"),4), axes = FALSE,

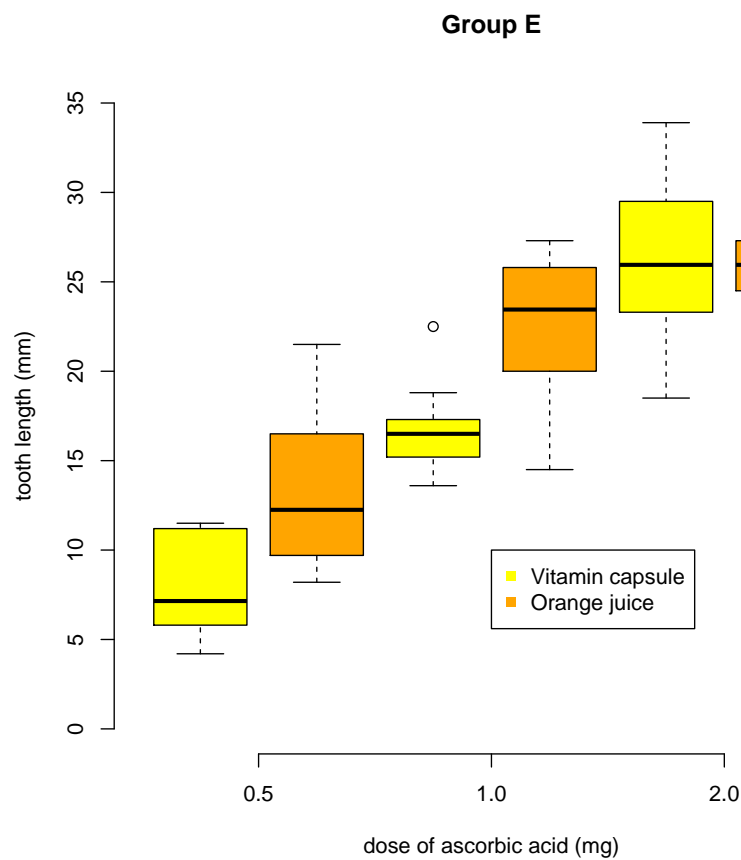
```



```

      xlab = "dose of ascorbic acid (mg)",
      ylab = "tooth length (mm)", ylim = c(0, 35),
      main = "Group E")
axis (2)
axis (1, label = c("0.5", "1.0", "2.0"),
      at = c(1.5, 3.5, 5.5))
legend(3.5, 10, pch = 15, col = c("yellow", "orange"),
      c("Vitamin capsule", "Orange juice"))

```



조F

supp의 레벨 처리를 하지 않았고 dose별 묶음 처리를 하지 않았다. 그래프의 전체 크기에 따라 바뀌지만 레전드의 위치도 재고하는 것이 좋겠다.

```

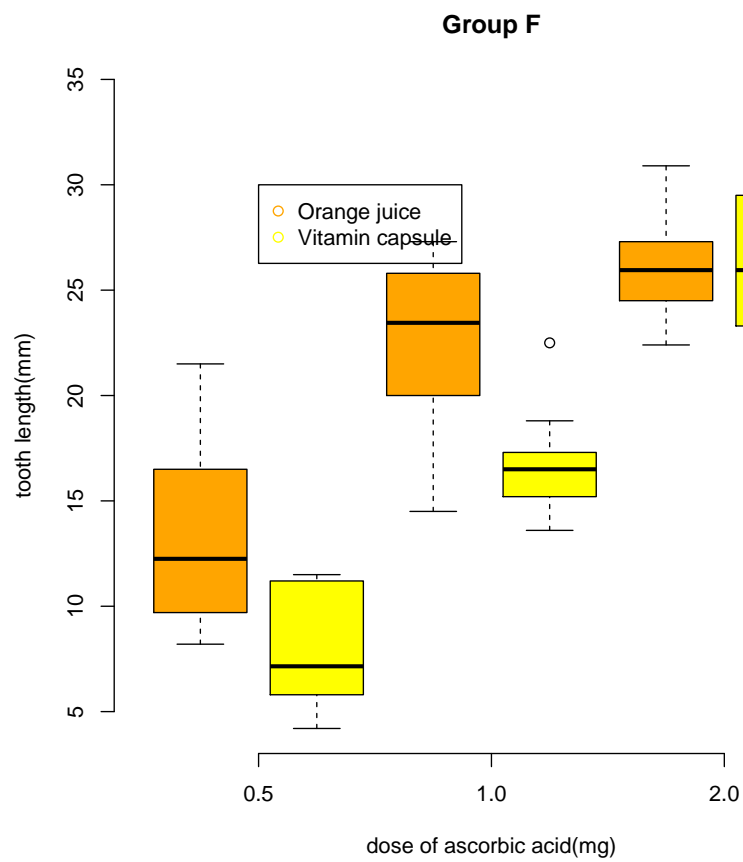
#-----
# 6조
# untidy levels
# failed grouping
# legend position

```

```

boxplot(len ~ supp*dose, ToothGrowth,
        main = "Group F",
        col= rep(c("orange", "yellow"),4),
        axes = FALSE,
        xlab= "dose of ascorbic acid(mg)", ylab="tooth length(mm)") #
axis(2)
axis(1, label = c("0.5", "1.0", "2.0"), at=c(1.5, 3.5, 5.5))
legend(1.5, 30, pch = 1, col = c("orange", "yellow"),
      c("Orange juice", "Vitamin capsule"))

```



이 문서는 knitr-X_yY_yTeX으로 작성되어 코드의 실행 결과는 모두 출력 당시에 실시간으로 이루어졌다. 문서는 R과 터미널에서 한 번씩 컴파일 해야 얻을 수 있다.

```
# in R
> library(knitr)
> knit("thisDocument.Rnw", output = "outputTex.tex")
.
.
# in terminal
$ xelatex outputTex.tex
```

컴파일에 필요한 컴퓨터의 조건으로 다음 어플리케이션들이 설치되어 있어야 한다: R, R에는 knitr 패키지가 필요하다. 운영체제에 맞는 pandoc, 그리고 TeXLive 최신 버전. 글꼴 Minion Pro, Myriad Pro, Courier가 Open Type Font 형태로 설치되어 있어야 컴파일 할 수 있다. 이 글꼴이 없다면 문서의 7부터 10행을 삭제한 뒤 컴파일 하면 된다.