#• 문제 1. ggplot2 의 midwest 데이터를 데이터 프레임 형태로 불러와서

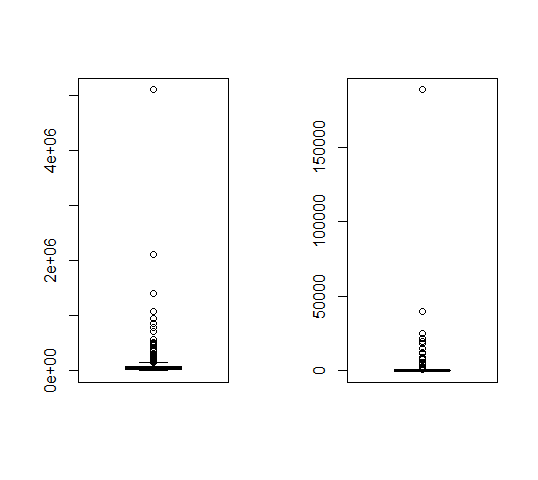
#poptotal과 popasian을 boxplot(), quantile()으로 데이터의 특성을 파악하세요.

midwest=as.data.frame(ggplot2::midwest)

midwest

boxplot(midwest$total)

boxplot(midwest$asian)



quantile(midwest$total)

quantile(midwest$asian)

> quantile(midwest$total)

0% 25% 50% 75% 100%

1701 18840 35324 75651 5105067

> quantile(midwest$asian)

0% 25% 50% 75% 100%

0 35 102 401 188565

#• 문제 2. poptotal(전체 인구)을 total 로, popasian(아시아 인구)을

#asian 으로 변수명을 수정하시오

midwest=rename(midwest,total=poptotal)

midwest=rename(midwest,asian=popasian)

#• 문제 3. total, asian 변수를 이용해 '전체 인구 대비 아시아 인구 백분율

# 파생변수를 만들고, 히스토그램과 확률밀도를 하나의 화면을

# 2개로 분활하여 만들어 도시들이 어떻게 분포하는지 살펴보세요.

midwest$asianper=100\*midwest$asian/midwest$total

frame()

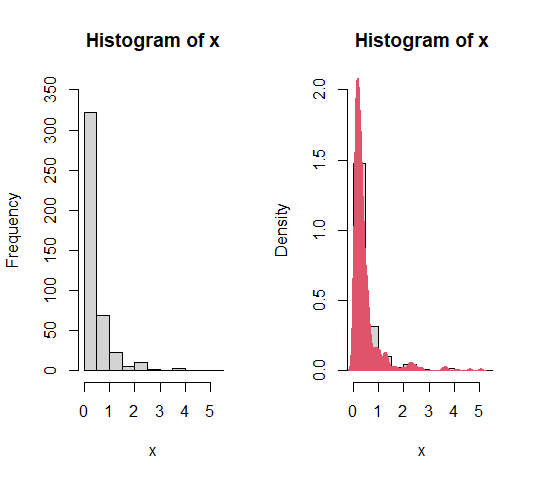
x=midwest$asianper

par(mfrow=c(1,2))

hist(x,ylim=c(0,350))

hist(x,,ylim=c(0,2),probability = T)

lines(density(x),col=2,type='h',lwd=2)



#• 문제 4. 아시아 인구 백분율 전체 평균을 구하고, 평균을 초과하면 "large",

# 그 외에는 "small"을 부여하는 파생변수를 만들어 보세요.

aspopav=mean(x)

> mean(x)

[1] 0.4872462

midwest$avgp=ifelse(x>0.4872462,"large","small")

#• 문제 5. "large"와 "small"에 해당하는 지역이 얼마나 되는지,

# 막대 그래프(barplot)를 만들어 확인해 보세요.

tt=table(midwest$avgp)

bb=barplot(tt,ylim=c(0,350),col=rainbow(3))

text(bb,tt,paste0(tt,"개"),pos=3,cex=2,col = 2)

