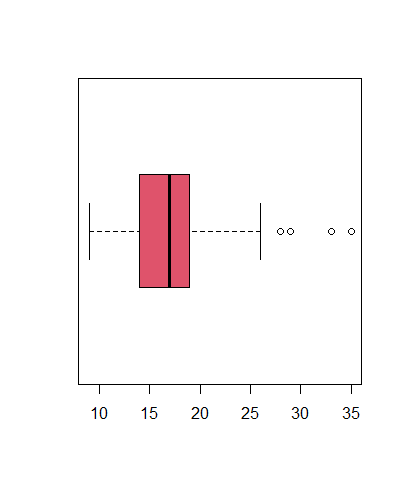
#4/29(금)\_HW1

1.mpg데이터내의 도시연비의 이상치를 NA값으로 처리하고

이를 제조사(manufacture)별로 평균값을 제시하시오?

mpg=data.frame(ggplot2::mpg)

HWB=boxplot(mpg$cty,horizontal = T,col=2)



HWB$stats

테이블이(가) 표시된 사진

자동 생성된 설명

mpg$cty=ifelse(mpg$cty>26|mpg$cty<9,NA,mpg$cty)

data.frame(mpg %>% group\_by(manufacturer) %>%

summarise(mean\_cty=mean(cty)))

텍스트이(가) 표시된 사진

자동 생성된 설명

2.mpg 데이터를 이용해서 분석 문제를 해결해 보세요.

우선 mpg 데이터를 불러와서 일부러 이상치를 만들겠습니다.

drv(구동방식) 변수의 값은 4(사륜구동),

f(전륜구동), r(후륜구동) 세 종류로 되어있습니다.

몇 개의 행에 존재할 수 없는 값 k를 할당하겠습니다.

cty(도시 연비) 변수도 몇 개의 행에 극단적으로 크거나 작은 값을 할당하겠습니다.

mpg <- as.data.frame(ggplot2::mpg) # mpg 데이터 불러오기

mpg[c(3,10, 14,20, 58, 93), "drv"] <- "k" # drv 이상치 할당

mpg[c(29, 43, 45, 48, 77, 129, 203), "cty"] <- c(3, 2, 4, 39, 42,47,49) # cty 이상치 할당

2.1. drv 에 이상치가 있는지 확인하세요. 이상치를 결측 처리한 다음 이상치가

사라졌는지 확인하세요. 결측처리 할 때는 %in% 기호를 활용하세요.

table(mpg$drv%in%c("4","f","r"))



mpg$drv=ifelse(mpg$drv%in%c("4","f","r"),NA,mpg$drv)

table(is.na(mpg$drv))

텍스트이(가) 표시된 사진

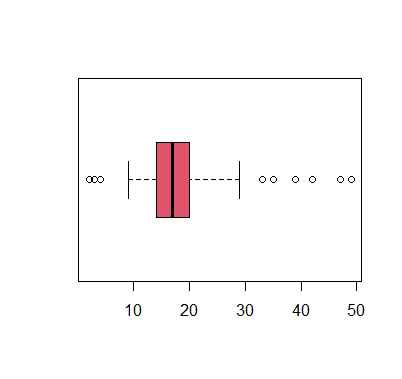
자동 생성된 설명

• 2.2. 상자 그림을 이용해서 cty 에 이상치가 있는지 확인하세요.

상자 그림의 통계치를 이용해 정상 범위를 벗어난 값을 결측 처리한 후 다시

상자 그림을 만들어 이상치가 사라졌는지 확인하세요.

HWB1=boxplot(mpg$cty,horizontal = T,col = 2)



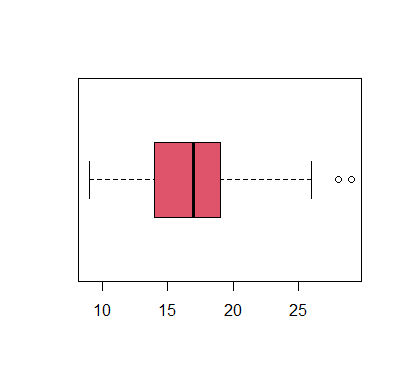
HWB1$stats’

테이블이(가) 표시된 사진

자동 생성된 설명

mpg$cty=ifelse(mpg$cty>29|mpg$cty<9,NA,mpg$cty)

HWB1=boxplot(mpg$cty,horizontal = T,col = 2)



• 2.3. 두 변수의 이상치를 결측처리 했으니 이제 분석할 차례입니다.

이상치를 제외한 다음 drv 별로 cty평균이 어떻게 다른지 알아보세요.

하나의 dplyr 구문으로 만들어야 합니다

data.frame(mpg %>% filter(!is.na(drv)&!is.na(cty)) %>%

group\_by(drv) %>%

summarise(mean\_cty=mean(cty)))

텍스트이(가) 표시된 사진

자동 생성된 설명