Question 1

nrow(subset(house, VAL == 24, select = VAL))

Question 3

dat <- read.xlsx("getdata-data-DATA.gov\_NGAP.xlsx", sheetIndex=1, rowIndex=c(18:23), colIndex=c(7:15))

sum(dat$Zip\*dat$Ext,na.rm=T)

Question 4

fileUrl <- "http://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Frestaurants.xml"

doc <- xmlTreeParse(fileUrl,useInternal=TRUE)

rootNode <- xmlRoot(doc)

xmlName(rootNode)

zips <- xpathSApply(rootNode[[1]][[1]], "//zipcode",xmlValue)

length(subset(zips, zips == 21231))

Question 5

> system.time( for(ii in 1:10000) DT[,mean(pwgtp15),by=SEX] )

user system elapsed

9.242 0.128 9.370

> system.time( for(ii in 1:10000) sapply(split(DT$pwgtp15,DT$SEX),mean) )

user system elapsed

8.222 1.092 9.313

> system.time( for(ii in 1:10000) tapply(DT$pwgtp15,DT$SEX,mean) )

user system elapsed

14.650 2.102 16.749