```
#Clean raw data from
https://www1.nyc.gov/site/finance/taxes/property-rolling-sales-data.page

library(readxl)
library(xlsx) #ingest library packages needed for coding

setwd("C:/Users/Rosario Household/Downloads") #set the working drectory to locate
the raw data from

rollingsales_nyc <- read_excel("rollingsales_nyc.xlsx") #ingest data into R

View(rollingsales_nyc) #View and study the data

summary(rollingsales_nyc) #Continuation of study of data

rollingnyc.sub1 <- subset(rollingsales_nyc, `SALE PRICE` > 0) # Remove any
Transactions without a sales price

rollingnyc.sub2 <- subset(rollingnyc.sub1, `BUILDING CLASS CATEGORY` ==         "01
ONE FAMILY DWELLINGS" |
                            `BUILDING CLASS CATEGORY` ==        "02 TWO FAMILY
DWELLINGS" |
                            `BUILDING CLASS CATEGORY` ==        "03 THREE FAMILY
DWELLINGS" |
                            `BUILDING CLASS CATEGORY` ==        "04 TAX CLASS 1
CONDOS" |
                            `BUILDING CLASS CATEGORY` ==        "12 CONDOS - WALKUP
APARTMENTS" |
                            `BUILDING CLASS CATEGORY` ==        "13 CONDOS -
ELEVATOR APARTMENTS")  #Filter required data for analysis

rollingnyc.sub2$monthyear <- format(as.Date(rollingnyc.sub2$`SALE DATE`),
"%m-01-%Y") #Format the date in the dataset to the first of corresponding month


rollingnyc.sub3 <- with(rollingnyc.sub2, aggregate(`SALE PRICE` ~ `ZIP CODE` +
monthyear, FUN =  function(x) c( md = median(x)) ) ) #Final output to obtain the
median per zip code and month

write.csv(rollingnyc.sub3, "C:/Users/Rosario Household/rl3.csv") #export the data
into a csv for tableau
```