*# Assignment 4 C*

[**rm**](http://inside-r.org/r-doc/base/rm)([**list**](http://inside-r.org/r-doc/base/list) = [**ls**](http://inside-r.org/r-doc/base/ls)())

*# installing and loading the XML package*

[**install.packages**](http://inside-r.org/r-doc/utils/install.packages)("XML")

[**library**](http://inside-r.org/r-doc/base/library)([XML](http://inside-r.org/packages/cran/XML))

*# Task 1*

*# Load and then parse the XML document at the URL*

*# Create any intermediate data objects as deemed necessary*

*# how many auctions had more than 5 bids*

*# parsing the XML document*

biddinGINfourl<-xmlInternalTreeParse("http://www.cs.washington.edu/research/xmldatasets/data/auctions/ebay.xml")

*# to get all the auction info and filter the auction with more than five bids*

numbeRINfo<-xpathApply(biddinGINfourl,"//listing/auction\_info[num\_bids>5]/num\_bids", [**function**](http://inside-r.org/r-doc/base/function)(n) [**as.numeric**](http://inside-r.org/r-doc/base/as.numeric)(xmlValue(n)))

*# print the number of bids*

bids<-[**length**](http://inside-r.org/r-doc/base/length)(numbeRINfo)

[**cat**](http://inside-r.org/r-doc/base/cat)("the number of auctions with more than five bids")

[**print**](http://inside-r.org/r-doc/base/print)(bids)

**strategy:**

* ***parsing the XML document***
* ***to get all the auction info and filter the auction with more than five bids through using xpathaply***
* ***print the number of bids***

*# TASK 2*

*# 2a) what was the highest closing price for the security?*

*# load the data and convert it into Data Frame for all the solutions below*

TradeDataUrl <-"http://www.barchartmarketdata.com/data-samples/getHistory15.xml"

TradeData <-xmlToDataFrame(TradeDataUrl)

*# funtion to get highest closing price for security*

HighestClosingPrice<- [**function**](http://inside-r.org/r-doc/base/function)(Price){

*# get the column for highest closing and print*

MaxPrice<-[**max**](http://inside-r.org/r-doc/base/max)(TradeData$high, na.rm = T)

[**cat**](http://inside-r.org/r-doc/base/cat)("highest closing price for the security")

[**cat**](http://inside-r.org/r-doc/base/cat)("**\n**")

[**print**](http://inside-r.org/r-doc/base/print)(MaxPrice)

}

HighestClosingPrice(ClosingPrice)

**strategy:**

* **define a function HighestClosingPrice**
* **get the required column from the dataframe**
* **filter the information and print the information**

*# 2b)what was the total volume traded?*

*# load the data and convert it into a dataframe*

*# write a function to get the total volume*

VolumeTraded <- [**function**](http://inside-r.org/r-doc/base/function)([**volume**](http://inside-r.org/r-doc/cluster/volume)){

*# convert the class as numeric and get the volume column and find the sum*

TotalVolume <- [**sum**](http://inside-r.org/r-doc/base/sum)(**[as.numeric](http://inside-r.org/r-doc/base/as.numeric)**(TradeData$volume), na.rm = T)

*# print the sum of the the total volume*

[**cat**](http://inside-r.org/r-doc/base/cat)(" the total volume traded")

[**cat**](http://inside-r.org/r-doc/base/cat)("**\n**")

[**print**](http://inside-r.org/r-doc/base/print)(TotalVolume)

}

VolumeTraded(VolumeTraded)

**Strategy:**

* **Write a function to get the total volume**
* **Get the required column, convert the class as numeric**
* **Get the sum of volume of trade**
* **Print the result**

*# 2c)what was the average trading volume during each HOUR of the trading day; place the result into a data frame*

*# function to extract the volume from the data frame*

HourAvg<-**[as.numeric](http://inside-r.org/r-doc/base/as.numeric)**(**[as.character](http://inside-r.org/r-doc/base/as.character)**(TradeData$volume))

*# removing the status*

HourAvg<-HourAvg[-1]

HourAvg<-[**head**](http://inside-r.org/r-doc/utils/head)(**[cbind](http://inside-r.org/r-doc/base/cbind)**(HourAvg), n= 92)

*# splitting the data into equal length*

HourAvg <- [**split**](http://inside-r.org/r-doc/base/split)(HourAvg, [**rep**](http://inside-r.org/r-doc/base/rep)(**[seq](http://inside-r.org/r-doc/base/seq)**(23), [**length**](http://inside-r.org/r-doc/base/length)(92), each=(4)))

HourAvg<-**[as.data.frame](http://inside-r.org/r-doc/base/as.data.frame)**(HourAvg)

*#finding the average of each column in the data frame*

**for**(i **in** 1:23)

{

Avg = [**sum**](http://inside-r.org/r-doc/base/sum)(HourAvg[i]/4)

[**print**](http://inside-r.org/r-doc/base/print)(Avg)

}

Strategy:

* To get the average trading volume during each hour we convert the data into a data frame
* Split the volume column into 23 columns of 4 rows each to divide them on hourly basis as the data is given on quarter hour basis
* Find the average of each column by applying for loop
* Print the results