IST-687 Chapter Notes Template: After Completing Please Submit as a PDF.
Originality Assertion: By submitting this file you affirm that this writing is your own.

Name: Yicun Deng
Class: M003
Date: 9/22/20
Chapter Number: #11
Title of Chapter: Storage Wars

Module 5 Lectures

We are going to focus on import data this week(different ways:excel JSON, database)
Two ways of format, flat and system-oriented data.
data() to see the embedded dataset for R
help()to see the actual dataset’s information(embedded only)
you can use Rstudio to directly read dataset
use gdata package
read.xls function

```
install.packages(“gdata”)
library(“gdata”)
testFrame<-read.xls()
testFrame<-testFrame[,1:5]
Numberize<-function(inputVector){
inputVector<-str_replace_all(inputVector,”,”,””)#remove comma
inputVector<-str_replace_all(inputVector,”,” ””)#remove space
return(as.numeric(inputVector))#return number
testFrame$newNames <- Numerize(testFrame$oldvectors)
```

Answers to the questions: When the excel files is built in another format of dataset, the readin function like this would not be appropriate.

Roundtable - Importing Data

The data is not always in a nice structured format like csv or sometimes in other nice structured format. That’s why we don’t always use read csv.
It doesn’t work when it has hierarchy on the database too.

Module 5 - SQL from R

RODBC package ust be loaded

```
ConSQL2012 <- odbcConnect(“SQL201264b”)
TblsSQL2012 <-sqlTables(conSQL2012)
DataSQL2012<-sqlQuery(conSQL2012,paste{”select*fromProducts”)}
DataSQL2012
View(DataSQL2012)

install.packages(“sqldf”)
library(“sqldf”)
sqldf(‘select mtcars.mpg from mtcars’)
```

IST-687 Chapter Notes Template: After Completing Please Submit as a PDF.
Originality Assertion: By submitting this file you affirm that this writing is your own.

```
sqldf(‘select AVG(mtcars.mpg) from mtcars where cyl=4’)#use sql code inside R
sapply(Variable,Function, optional parameters
tapply(summary Variable, group Variable, Function)#apply the function to the summary variable
through group variable
```

Module 5 - JSON

JSON is a webbased standard of access server’s information

```
MakeGeoURL <- function(address){
Root <- http://maps.google.com/maps/api/geocode/
url<- paste(root, “json?address=”,address,”&sensor=false”, sep=””)
return(URLencode(url))
}
Addr2latlng<-function(address){
url<-MakeGeoURL(address)
apiResult<-getURL(url)
geoStruct <- fromJSON(apiResult,simplify = FALSE)
lat<-NA
lng<-NA
try(lat<-geoStruct$results[[1]]$geometry$location$lat)
try(lat<-geoStruct$results[[1]]$geometry$location$lng)
return(c(lat,lng))
}

numRows<-length(stations)
nameList<-names(stations[[1]])

dfstations <-data.frame(matrix(unlist(stations),
nrows=numRows, byrow=T),
stringsAsFactors=FALSE)
names(dfStations) <-nameList
```

Answer: They want people to be able to do analyze on their data hence give them an idea of the improving their services. Bad side is that it might show what they are lacking of to the public if any.

Open and sharing is the purpose of opening JSON, it is human readable too(friendly to us haha)

Data Science Roundtable 3

Airplane bumping would give you some compensation for the bumped customer, but it is still bad experience. Using data science on this could reduce the possibility of flight bumping and make it both easier for airline to run and customer to enjoy a better flight experience instead of unnecessary waiting time.

IST-687 Chapter Notes Template: After Completing Please Submit as a PDF.
Originality Assertion: By submitting this file you affirm that this writing is your own.

Exercise Review

```
install.packages("RCurl")
install.packages("RJSONIO")
library("RCurl")
library("RJSONIO")
```

```
parkURL <- "https://parknyc.parkmobile.us/PhonixxAPI/json/metadata?op=ParkingZone"
apiResult<-getURL(parkURL)
results<-fromJSON(apiResult)
length(results)
```

error messages:

```
Error in function (type, msg, asError = TRUE)  : 
  error:1407742E:SSL routines:SSL23_GET_SERVER_HELLO:tlsv1 alert protocol
> results<-fromJSON(apiResult)
Error in fromJSON(apiResult) : object 'apiResult' not found
> length(results)
Error: object 'results' not found

>
```

Question for Class

For my homework4

I used this code

```
vectorStats <- function(x){#recreating the function vectorStats
  result <- c(mean(x),median(x),max(x),min(x),sd(x))#get the mean median max min sd by order
  return(result)#return the result
}
```

And it returns a vector, I want to know how to make all these vectors returned after the repeating of samples into a dataframe.. When I try to modify the functions and codes, the best I got is a dataframe with 5 variables and 1000obs….

Also, I want to know why does my RStudio gave me these error message after I tried the code for the challenge… It is because of the warning that they are built for different RStudio edition?

Sorry for the amount of questions…