Project

q.Part I

There are 20 files with .dat extention. You have to read all the files in to single dataframe.

a.1

First set the directory to location of 20 files

setwd("C:\\Users\\Downloads\\iris\_dat\_file\\iris")

Read all the files using function and read.delim

f\_read<- function(path1) {

path1="C:\\Users\\Chirag Bhawna\\Downloads\\iris\_dat\_file\\iris"

filenames= list.files(path=path2, full.names = T)

for(i in 1){

f[i]=read.delim(filenames[i],header=F, sep=",", skip=9)

f[i+1]=read.delim(filenames[i+1],header=F, sep=",", skip=9)

f[i+2]=read.delim(filenames[i+2],header=F, sep=",", skip=9)

f[i+3]=read.delim(filenames[i+3],header=F, sep=",", skip=9)

f[i+4]=read.delim(filenames[i+4],header=F, sep=",", skip=9)

f[i+5]=read.delim(filenames[i+5],header=F, sep=",", skip=9)

f[i+6]=read.delim(filenames[i+6],header=F, sep=",", skip=9)

f[i+7]=read.delim(filenames[i+7],header=F, sep=",", skip=9)

f[i+8]=read.delim(filenames[i+8],header=F, sep=",", skip=9)

f[i+9]=read.delim(filenames[i+9],header=F, sep=",", skip=9)

f[i+10]=read.delim(filenames[i+10],header=F, sep=",", skip=9)

f[i+11]=read.delim(filenames[i+11],header=F, sep=",", skip=9)

f[i+12]=read.delim(filenames[i+12],header=F, sep=",", skip=9)

f[i+13]=read.delim(filenames[i+13],header=F, sep=",", skip=9)

f[i+14]=read.delim(filenames[i+14],header=F, sep=",", skip=9)

f[i+15]=read.delim(filenames[i+15],header=F, sep=",", skip=9)

f[i+16]=read.delim(filenames[i+16],header=F, sep=",", skip=9)

f[i+17]=read.delim(filenames[i+17],header=F, sep=",", skip=9)

f[i+18]=read.delim(filenames[i+18],header=F, sep=",", skip=9)

f[i+19]=read.delim(filenames[i+19],header=F, sep=",", skip=9)

f[i+20]=read.delim(filenames[i+20],header=F, sep=",", skip=9)

mergedata<- rbind(f[i],f[i+1],f[i+2],f[i+3],f[i+4],f[i+5],f[i+6],f[i+7],f[i+8],f[i+9],f[i+10],f[i+11],f[i+12],f[i+13],f[i+14],f[i+15],f[i+16],f[i+17],f[i+18],f[i+19],f[i+20])

return(mergedata)

}

}

Part II

The data is present in xml format, with file name, iris.xml. Your task is to read the XML data and store it

in the data frame df.

A.

To read XML data we first need to install XML package.

install.packages("xml")

Installing package into ‘C:/Users/Chirag Bhawna/Documents/R/win-library/3.4’

(as ‘lib’ is unspecified)

Warning in install.packages :

package ‘xml’ is not available (for R version 3.4.2)

Warning in install.packages :

Perhaps you meant ‘XML’ ?

> library(XML)

> xml\_data<- xmlParse(file = "books.xml")

xml\_data<- xmlParse(file = "books.xml")

> xml\_data

<?xml version="1.0"?>

<catalog>

<book id="bk101">

<author>Gambardella, Matthew</author>

<title>XML Developer's Guide</title>

<genre>Computer</genre>

<price>44.95</price>

<publish\_date>2000-10-01</publish\_date>

<description>An in-depth look at creating applications

with XML.</description>

</book>

<book id="bk102">

<author>Ralls, Kim</author>

<title>Midnight Rain</title>

<genre>Fantasy</genre>

<price>5.95</price>

<publish\_date>2000-12-16</publish\_date>

<description>A former architect battles corporate zombies,

an evil sorceress, and her own childhood to become queen

of the world.</description>

</book>

<book id="bk103">

<author>Corets, Eva</author>

<title>Maeve Ascendant</title>

<genre>Fantasy</genre>

<price>5.95</price>

<publish\_date>2000-11-17</publish\_date>

<description>After the collapse of a nanotechnology

society in England, the young survivors lay the

foundation for a new society.</description>

</book>

Part III

Convert the iris data into the JSON format and read the data in JSON format and convert it into

dataframe “iris\_data”.

A.

To read json data in R we need to install package “rjson”

install.packages("rjson")

> read\_json<-fromJSON(file = "one.json")

>

> read\_json

$ID

[1] "1" "2" "3" "4" "5" "6" "7" "8"

$Name

[1] "Rick" "Dan" "Michelle" "Ryan" "Gary" "Nina" "Simon"

[8] "Guru"

$Salary

[1] "623.3" "515.2" "611" "729" "843.25" "578" "632.8" "722.5"

$StartDate

[1] "1/1/2012" "9/23/2013" "11/15/2014" "5/11/2014" "3/27/2015" "5/21/2013"

[7] "7/30/2013" "6/17/2014"

$Dept

[1] "IT" "Operations" "IT" "HR" "Finance" "IT"

[7] "Operations" "Finance"

> iris\_json<-toJSON(iris)

> iris\_json

[1] "{\"Sepal.Length\":[5.1,4.9,4.7,4.6,5,5.4,4.6,5,4.4,4.9,5.4,4.8,4.8,4.3,5.8,5.7,5.4,5.1,5.7,5.1,5.4,5.1,4.6,5.1,4.8,5,5,5.2,5.2,4.7,4.8,5.4,5.2,5.5,4.9,5,5.5,4.9,4.4,5.1,5,4.5,4.4,5,5.1,4.8,5.1,4.6,5.3,5,7,6.4,6.9,5.5,6.5,5.7,6.3,4.9,6.6,5.2,5,5.9,6,6.1,5.6,6.7,5.6,5.8,6.2,5.6,5.9,6.1,6.3,6.1,6.4,6.6,6.8,6.7,6,5.7,5.5,5.5,5.8,6,5.4,6,6.7,6.3,5.6,5.5,5.5,6.1,5.8,5,5.6,5.7,5.7,6.2,5.1,5.7,6.3,5.8,7.1,6.3,6.5,7.6,4.9,7.3,6.7,7.2,6.5,6.4,6.8,5.7,5.8,6.4,6.5,7.7,7.7,6,6.9,5.6,7.7,6.3,6.7,7.2,6.2,6.1,6.4,7.2,7.4,7.9,6.4,6.3,6.1,7.7,6.3,6.4,6,6.9,6.7,6.9,5.8,6.8,6.7,6.7,6.3,6.5,6.2,5.9],\"Sepal.Width\":[3.5,3,3.2,3.1,3.6,3.9,3.4,3.4,2.9,3.1,3.7,3.4,3,3,4,4.4,3.9,3.5,3.8,3.8,3.4,3.7,3.6,3.3,3.4,3,3.4,3.5,3.4,3.2,3.1,3.4,4.1,4.2,3.1,3.2,3.5,3.6,3,3.4,3.5,2.3,3.2,3.5,3.8,3,3.8,3.2,3.7,3.3,3.2,3.2,3.1,2.3,2.8,2.8,3.3,2.4,2.9,2.7,2,3,2.2,2.9,2.9,3.1,3,2.7,2.2,2.5,3.2,2.8,2.5,2.8,2.9,3,2.8,3,2.9,2.6,2.4,2.4,2.7,2.7,3,3.4,3.1,2.3,3,2.5,2.6,3,2.6,2.3,2.7,3,2.9,2.9,2.5,2.8,3.3,2.7,3,2.9,3,3,2.5,2.... <truncated>

---To convert json data into data frame we use following command

json\_data\_frame <- as.data.frame(read\_json)

Part IV

Use dplyr function on the data iris\_data. Implement select, match, filter, arrange, rename, and mutate

function on the iris\_data.

A.

Iris\_dplyr <-iris

Iris\_dplyr %>% select(sepal.length, petal. Length, species) %>% group\_by(species)%>% summarise(mean\_length=mean(sepal.length)) >%> mutate(mean\_length) %>% filter(mean\_length > 6.0)

Part V

Print summary

Summary(Iris\_dplyr)