R in Windows: Startup on ITaP mounted files

Click start on lower left in search box: [R] [return]

Click on RProjectx64 3.1.1

The R working directory (i.e., the folder with .RData and often files read and written)

Get the name of the current working directory

```
> getwd()
```

```
[1] "\\\\myhome.itap.purdue.edu/puhome/My Documents"
```

This is the ITaP default

Change this working directory, first make folder W:/R_course outside of R.

```
> setwd("W:/R_course")
```

```
> getwd()
```

```
[1] "W:/R_course"
```

List files in the current working directory

```
> list.files()
```

R in Linux or iOS: Startup

Open a Terminal

`cd` into the directory which will be the working directory

Then type R

```
> getwd()
> setwd("/home/tongx/R_course")
> list.files()
```

Login to Linux server on Windows machine

Active the connection between local machine and server

Click the Start on lower left in search box: [putty] [return]

Click on putty.63

In the Host Name, type in: `hathi.rcac.purdue.edu`. Port keep as 22. Connection type is: `SSH`.

In Saved Sessions, type in `Hathi`.

Then click the Save button on the right.

Type in your Purdue account and the password. Then press Enter.

Now you are successfully connecting to the front-end of Hathi cluster.

Login to Linux Server on Linux Machine

Open an command terminal

Type in: `ssh yourusername@hathi.rcac.purdue.edu`

Type in your passward.

Now you are on Hathi front-end

Hathi Cluster

One front-end server into which user logs in

- 16 cores
- 24.5 GG RAM

6 nodes on the Hadoop cluster

Each node

- 16 cores
- 32 GB memory
- HDFS storage: 48 TB

Hathi Cluster

Try some linux commands:

`mkdir R_LIBS` : create a directory named R_LIBS

`mkdir stat695v` : create a directory named stat695v

`ls` : list all files and directories in current working directory

`pwd` : print current working directory

`cd stat695v` : change current working direcotry to stat695v

```
module add r/3.1.0
```

Hathi cluster

Type `R` in the terminal. Notice that the R you start will not be on your laptop, it is on Hathi front-end node.

```
> getwd()
> library(lattice)
> x <- 1:10
> y <- rnorm(10)
> trellis.device( device = pdf, file = "plot1.pdf")
> plot(x,y)
> dev.off()
```

After this, let's quite R by typing `q()` in R. Return to the Linux command terminal, check if PDF file has been created in your current working directory.

```
ls
```

Files management

We need a method to visualize the graph which is not highly depends on the bandwith of internet.

On Linux laptop, `scp` command can be used to copy files from cluster to the local machine.

```
scp
tongx@hathi.rcac.purdue.edu:/home/tongx/plot1.pdf
/home/tongx
```

Files management

On Windows laptop, we have to come up with a different way.

First click ”Start” on lower left, click on ”Computer”.

Click right butten of your mouse on the ”Computer” of left side, then click on Map network dirve...

In the new popping up window, in the Drive session, choose X:

In the Folder session, type in:

\\samba.rcac.purdue.edu\yourusername

Also check the ”Reconnect at logon” box

Then click Finish. Now you should be able to see your home directory on Hathi in the Network session of Computer.

You can move, copy, edit files you owned on Hathi.

Editing R scripts

Once login to the Hathi front-end, you can edit R scripts by using text editor like vi, vim, emacs.

For Windows laptop, you still can use Windows text editor like wordtext to edit R script files. But before you source those script files in R, you have to run following command in the command terminal on Hathi front-end node.

```
dos2unix filename.R
```

This linux command will help us remove those special characters that created by Windows text editor but cannot be recognized by Linux system.

Performance Test: Hathi Front End

First download the psl.csv dataset from course webpage to your local Windows machine.

Recall that we have mapped our home directory on Hathi front-end to Network dirve on your local Windows machine. Copy the `psl.csv` from your local Windows machine to your Hathi home directory.

Then login to Hathi front-end through Putty. Or ssh if you are using a Linux/iOS laptop.

Type `R` in your command line to start R. Make sure `psl.csv` is in your R current working directory.

```
list.files()
```

Performance Test: Hathi Front End

Read in dataset into R:

```
> psl.df <- read.csv("psl.csv")
```

Get information of the dataset:

```
> dim(psl.df)
```

```
> head(psl.df)
```

Fit linear regression:

```
> lm(log2(perseat) ~ row, data = psl.df)
```

Timing the fitting procedure:

```
> system.time(lm(log2(perseat) ~ row, data =
psl.df))
```

Performance Test: Hathi Front End

Let us double the size of dataset:

```
> psl.new <- rbind(psl.df, psl.df)
```

Timing the fitting procedure:

```
> system.time(lm(log2(perseat) ~ row, data =
psl.new))
```

Double the size again:

```
> psl.new <- rbind(psl.new, psl.df)

> object.size(psl.new)

> system.time(lm(log2(perseat) ~ row, data =
psl.new))
```