Introduction to Data Science

Live Session 07 - Unit 07

Future Plan

- Feb 28 : Live session 07 – unit 07
- Mar 07 : Live session 08 – Unit 08 (Case Study 1)
- Mar 14: Live session 09 – Unit 09 (API presentations)

Case study 01

- Feb 28 –
- Mar 07 – Discussion (Case study 01)
- Mar 14 – Submit before the live session

API Presentations (9.3)

- Install and load one of the packages given in the list for downloading APIs on this link: https://github.com/ropensci/opendata . The video gives a different URL, but I think this one is easier to navigate.
- The link is to a GitHub page. Examine the README file to find a list of R functions that interface with various APIs.
- Choose your favorite subject and select ONE R package from the list.
- Find/create an example to download some data using that library.
- Change an argument to the function/example to see what it does.
- Create a PPT presentation to show in the live session.
- And above all - have fun with it!

Future Plan

- Mar 14 – Unit 9 Videos (API presentations)
- Mar 21 – Unit 10
- Mar 28 – No live session
- Apr 04 – Unit 11
- Apr 11 – Unit 12 Videos (Python)
- April 17 – Python Presentations
- April 24 – Case Study II

Office hours

- Raunak : Friday 6.30p.m.-7.30p.m. CT
- Chen Mo : Friday 8.30p.m. – 9.30p.m. CT

Objectives

- Importing locally stored data sets
- Importing data from non-secure URLs
- Importing compressed data
- Regular Expressions in R
- Principles of Tidy Data
- Cleaning Data for Merging
- Recoding Variables
- Merging Data

Discussion

WHY DO WE CARE ABOUT TIDY DATA?

Importing Data from Non-secure URLs

- `site = "http://www.users.miamioh.edu/hughesmr/sta333/"`
- Pick a data set (eg: baseballsalaries)
- Import into R
- Tell me how many variables and observations
- What are the variables?

```
library(repmis)

site =
"http://www.users.miamioh.edu/hughesmr/sta333/baseballsalaries
.txt"
download.file(site,destfile= "./baseballsalaries.txt")
list.files()
#baseball<-read.table(file.choose(),header=TRUE)
baseball<-read.table("baseballsalaries.txt",header=TRUE)
head(baseball)
dim(baseball)
names(baseball)
```

Breakout Assignment

- In breakout rooms, you will be assigned two of the eight functions documented in the R package repmis
  - First, define the function
  - Second, give an example with R code (hopefully a working one)
  - Exception: InstallOldPackages.
- You can use the example in the help file.
- Designate a group member to present the findings.

repmis

- Miscellaneous tools for reproducible research
- Functions
  - git_stamp
  - InstallOldPackages
  - LoadandCite
  - scan_https
  - set_valid_wd
  - source_data
  - source_DropboxData (no longer supported)
  - source_XlsxData

Helpful Links

- http://www.r-bloggers.com/this-r-data-import-tutorial-is-everything-you-need/
- https://www.datacamp.com/community/tutorials/r-tutorial-read-excel-into-r#gs.qdlV1mw
- http://www.statmethods.net/input/importingdata.html