DataSci Cleaning Data Lecture Notes: Downloading Files

Using R to download files: getting and setting your directions

1. Basic component to working with data is knowing the working directory
2. Getwd() and setwd() are the main commands
3. Relative file path: setwd(“./data”), or setwd(“../”)
   1. Setwd(“../”) Moves it up one director
4. Absolute file paty: setwd(“/Users/jtleek/data/”)
5. Windows filepaths: setwd(“C:\\Users\\Andrew\\Downloads”)
   1. Use backslashes instead of forwards slashes on a machine

Checking for and Creating Directories

1. File.exists(“directoryName”) checks to see if the directory exists
2. Dir.create(“directoryName”) will create a directory if it does not exist
3. Example
   1. If(!file.exists(“data”)){

Dir.create(“data”)

}

Getting data from the internet

1. Use download.file() command to download from the internet
2. Yes, it can be done by hand but this insures reproducibility
3. Uses the parameters *URL, destfile, method* 
   1. URL – the URL of the data
   2. Destfile – destination file for where it goes
   3. Method - ???
4. Useful for tab-deliminated, csv, etc

Example: Baltimore Fixed Camera data

1. Has an export file (such as a csv)
2. Right click and that lets you the copy download link URL
3. Use that as your URL parameter in download.file
4. For dest files, use the extension associated with the file, (.csv for comma deliminated)
5. Method = “curl” for any https URL link for MACs
6. Check the date() command for the file to verify the download date and thus infer the most current version of the file

Download.file() info

1. Method can be left empty for http and some https
2. Large files take awhile to download
3. Record when the file was downloaded.