MFE R Programming Workshop

Week 5

Mahyar Kargar and Dan Yavorsky

Fall 2018

Introduction

Questions

Any questions before we start?

Overview

- Importing Data from Databases
- Importing Data from the Web
- Importing Data from WRDS
- `data.table`

Importing Data from Databases

Databases

- A database stores data in table format.
- There a several popular database management systems (DMS):
  - MySQL, PostgreSQL, SQLite, Oracle, Microsoft SQL Server, etc.
- Structured Query Language (SQL) is used maintain and query the database.

Databases in R

- There are different R packages for each DMS:
  - MySQL = RMySQL
  - PostgreSQL = RPostgreSQL
  - SQLite = RSQLite
  - Oracle = ROracle
  - SQL Server = RSQLServer
- The functions we use to interact with the database are specified in the R package `DBI`.

Importing Data From the Web

Quandl

- Quandl is a useful source of financial data and there is an R package `Quandl` to import the data into R.
- See https://www.quandl.com/tools/r.
- Data can be downloaded as `xts` objects, datatables, etc.

```
library(Quandl)
# download GDP % growth as an xts object
gdp <- Quandl("FRED/GDP", type="xts")
last(gdp, 4)
```

```
##                [,1]
## 2017 Q2 19250.01
## 2017 Q3 19500.60
## 2017 Q4 19754.10
## 2018 Q1 19956.81
```

Importing Data from WRDS

WRDS, CRSP, and R

- Wharton Research Data Services (wrds) has over 250 terabytes of data.
- One data provider is The Center for Research in Security Prices (CRSP).
  - You will use CRSP data throughout the MFE program.
- I will show you how to access WRDS from R.
- Documentation: Using R with WRDS

Setup

- First, you need to encode your wrds password: instructions here.
- We also need to obtain access to WRDS and download the SAS drivers for JDBC from here.
- The two files should be saved locally.
- Take note of the path to the files; we need the path to

data.table

What is a data.table?

- Think of data.table as an advanced version of data.frame.
  - Every column is the same length, but may have a different type
- It inherits from data.frame and works perfectly even when data.frame syntax is applied on data.table.
- data.table is very fast.
- The syntax of data.table is very concise.
  - Lowers programmer time...
  - ...but it can be hard to understand
  - Make sure you comment your code!
- Highly recommend going through data.table Cheat Sheet.

```
library(data.table)
```

An Example

- Syntax is DT[i, j, by]