

Learning `data.table` Package for Finance

Neeraj Jain

```
library(data.table)
library(magrittr) #for pipe operations (just used once)
```

1. Loading R Libraries

```
price <- fread("https://raw.githubusercontent.com/Neeraj2308/DataSet/main/price.csv")
```

2. Reading data from GitHub Note: `fread` automatically detect date variable, and how values are separated. No need to add additional arguments.

Source: *Prowess*

```
price[, unique(company_name)] %>% length
```

3. Number of companies in the dataset

```
## [1] 10
```

Pipe operations make the coding simple by using `%>%`

```
indicators <- c("company_name", "co_stkdate", "bse_closing_price", "bse_returns", "bse_market_cap")
```

4. Variables to be considered Getting data for above variables only

```
price <- price[, ..indicators]
```

```
setorder(price, company_name, co_stkdate)
```

5. Easy to put data into order use `-` to put data into descending order..

```
setorder(price, -company_name, co_stkdate)
```

6. Creating new variables For example: taking return in decimal forms

```
price[, returns.d := bse_returns / 100]
```

```
price <- price[!is.na(returns.d)]
```

7. Removing NA values

Group Operations

Easy to performs operations group wise in data.table package

```
price[, .N, by = company_name]
```

8. Number of observations for each company

```
##               company_name      N
## 1:      EICHER MOTORS LTD.  7392
## 2: DR. REDDY'S LABORATORIES LTD. 7483
## 3:      DIVI'S LABORATORIES LTD. 4489
## 4:      COAL INDIA LTD.  2579
## 5:      CIPLA LTD.  7446
## 6:      BRITANNIA INDUSTRIES LTD. 7363
## 7:      BHARTI AIRTEL LTD.  4755
## 8:  BHARAT PETROLEUM CORPN. LTD. 6801
## 9:      BAJAJ FINANCE LTD.  6513
## 10:     ASIAN PAINTS LTD.  7475
```

```
price[, logret := c(NA, diff(log(bse_closing_price))), by = company_name]
```

#alternative using shift

#price[, logre := log(bse_closing_price / shift(bse_closing_price, 1)), by = company_name]

9. Calculating log returns own for each company

```
price <- price[, .SD[-1], by = company_name]
```

10. Removing first observation of each company, as return is NA

```
price[, .(Avg = mean(logret, na.rm = TRUE)*100,
             SD = sd(logret, na.rm = TRUE)*100,
             Min = min(logret, na.rm = TRUE)*100 ),
       by = company_name]
```

11. Getting average return, sd, min, max, etc for each company

```
##               company_name      Avg      SD      Min
## 1:      EICHER MOTORS LTD.  0.07070289  4.623578 -230.02948
## 2: DR. REDDY'S LABORATORIES LTD.  0.05820436  3.084957  -95.20088
## 3:      DIVI'S LABORATORIES LTD.  0.06740415  3.678582 -158.74232
## 4:      COAL INDIA LTD. -0.03744092  1.853884  -15.66800
```

```
## 5:          CIPLA LTD. -0.01794040 4.821989 -229.61126
## 6:    BRITANNIA INDUSTRIES LTD.  0.04718162 2.843062 -147.95627
## 7:          BHARTI AIRTEL LTD.  0.05176757 2.579431  -67.23549
## 8:  BHARAT PETROLEUM CORPN. LTD. -0.01248160 3.433130 -109.86123
## 9:          BAJAJ FINANCE LTD.  0.05026933 4.090252 -228.30446
## 10:         ASIAN PAINTS LTD.  0.03154937 3.520004 -230.50905
```

```
price[, .(stdev = sd(logret, na.rm = TRUE)),
        keyby = .(company_name, year(co_stkdate))]
```

12. Calculating standard deviation of returns for each company for each year .

```
##          company_name year      stdev
## 1:  ASIAN PAINTS LTD. 1990 0.02130631
## 2:  ASIAN PAINTS LTD. 1991 0.01783447
## 3:  ASIAN PAINTS LTD. 1992 0.04783160
## 4:  ASIAN PAINTS LTD. 1993 0.02437176
## 5:  ASIAN PAINTS LTD. 1994 0.03489028
## ---
## 265: EICHER MOTORS LTD. 2017 0.01439050
## 266: EICHER MOTORS LTD. 2018 0.01940471
## 267: EICHER MOTORS LTD. 2019 0.02373352
## 268: EICHER MOTORS LTD. 2020 0.14744166
## 269: EICHER MOTORS LTD. 2021 0.01832763
```

Note: keyby also put data into ascending order too. by can also be used.

13. Doing winsorization for each company (for logret only) Defining winsorization function

```
winsorize <- function(x, prob = .01) {
  q <- quantile(x, probs = c(prob, 1 - prob))
  x[x < q[1]] <- q[1]
  x[x > q[2]] <- q[2]
  return(x)
}
```

Doing winsorization:

```
price[, logret.w := winsorize(logret, prob = 0.01), by = company_name]
```

New variable created logret.w

```
summary(price[, .(logret, logret.w)])
```

14. comparing summary of winsorized data with non-winsorized data.

```
##          logret          logret.w
## Min.   :-2.3050905  Min.   :-0.0984401
## 1st Qu.: -0.0109860  1st Qu.: -0.0109860
## Median : 0.0000000   Median : 0.0000000
## Mean    : 0.0003375   Mean    : 0.0007869
## 3rd Qu.: 0.0118345   3rd Qu.: 0.0118345
## Max.    : 0.5738004   Max.    : 0.1112256
```

```
price.m <- price[, .SD[.N], keyby = .(company_name, year(co_stkdate), month(co_stkdate))]
```

15. getting month end price only or monthly data This can be also done by other packages like quantmod

```
price[, index.ret := mean(logret),  
      by = .(year(co_stkdate), yday(co_stkdate))]
```

16. Getting average return for each day and setting this as a variable Similarly we can calculate weighted return using market cap data. This can be done for sector wise also.

Final Note

we can use this package to do any kind of calculations required with financial data. Most important is working with this package is very fast.