

투자론

- R과 Excel을 통한 금융데이터 분석 -

8주차

R의 Collecting Data 및 Portfolio Analysis

충남대학교
장호규 교수

Unit 02

Portfolio Analysis with R (1)

Overview

- Data Management for Portfolio Analysis

◆ DataGuide

- We can download lots of financial data such as price, returns, financial statements, and etc from the publicly available websites.
- For instance, kind.krx.co.kr, finance.naver.com, data.krx.co.kr, wiseindex.com, and etc.
- But, we do not further explore into how to obtain financial data using API (crawl).
- Instead, we deal with financial data, especially price and return data, from DataGuide.

◆ DataGuide



The screenshot shows the DataGuide website with a dark blue header and a light blue main content area. The header includes the 'DataGuide' logo, navigation links (INTRODUCE, EDUCATION, USER GUIDE, CUSTOMER CENTER), and a language selector (English | Korean). The main banner features the text 'Your Best Financial Partner' and describes it as 'The ultimate set of financial analytic program for an expert of financial professionals. It places the massive financial market data and analytic tools.' Below the banner is a row of icons for Reports, Data Center, Tools, Refresh, Lookup, and Help. A computer monitor displays the DataGuide interface. Below the banner, there is a section titled '에프앤가이드가 제공하는 국내 최고의 투자분석 시스템' (The best investment analysis system provided by F&N Guide) with a 'Download' button. The bottom section is divided into three columns: 'INTRODUCE' (DataGuide introduction), 'EDUCATION' (DataGuide education), and 'USER GUIDE' (DataGuide user guide). Each column has a 'More »' link.

DataGuide

English | Korean

INTRODUCE EDUCATION USER GUIDE CUSTOMER CENTER

Your Best Financial Partner

The ultimate set of financial analytic program for an expert of financial professionals.
It places the massive financial market data and analytic tools.

Reports Data Center Tools Refresh Lookup Help

에프앤가이드가 제공하는 국내 최고의 투자분석 시스템
2016년 6월 현재 국내 260여개의 기관과 연구소 및 대학교에서 사용하고 있는 국내 최고의 데이터베이스 어플리케이션입니다.

프로그램 인출 및 설치
Download

INTRODUCE

데이터가이드 소개 [More »](#)

데이터가이드를 통해 금융시장 및 기업 분석에 필요한 데이터를 다양한 방식으로 조회가 가능하며, 다양하게 구성된 템플릿을 통해 원클릭으로 편리하게 데이터를 분석하실 수 있습니다. 또한 사용자의 Needs에 맞게 커스터마

EDUCATION

데이터가이드 교육 [More »](#)

데이터가이드 사용자들을 위한 교육을 격주로 실시하고 있습니다. 기본적인 사용법과 데이터에 대한 이해를 돕기 위한 일반 교육 과정을 제공합니다.

USER GUIDE

데이터가이드 활용 [More »](#)

DataGuide의 기본적인 사용법에 관한 Video Manual 및 활용방안에 관한 Case Studies를 통해 고객들이 궁금해하는 사항에 관하여 서비스합니다.

◆ DataGuide Company Financial Data

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Refresh	Last Refresh: 2022-09-21 16:27:56											
2	Calendar Basis												
3	Portfolio	Default											
4	Item												
5	Frequency	월간	KRW										
6	Non-Trading Date	Previous	Asc										
7	Include Weekends	ALL											
8	Term	20100101 Current(20220921)											
9	Symbol	A005930	A005930	A373220	A373220	A000660	A000660	A207940	A207940	A051910	A051910	A006400	A006400
10	Symbol Name	삼성전자	삼성전자	네지솔루션	네지솔루션	SK하이닉스	SK하이닉스	이오로지스	이오로지스	LG화학	LG화학	삼성SDI	삼성SDI
11	Kind	SSC	SSC	SSC	SSC	SSC	SSC	SSC	SSC	SSC	SSC	SSC	SSC
12	Item	S41000060F	S41000170F	S41000060F	S41000170F	S41000060F	S41000170F	S41000060F	S41000170F	S41000060F	S41000170F	S41000060F	S41000170F
13	Item Name	증가(원)	수익률(%)	증가(원)	수익률(%)	증가(원)	수익률(%)	증가(원)	수익률(%)	증가(원)	수익률(%)	증가(원)	수익률(%)
14	Frequency	MONTHLY	MONTHLY	MONTHLY	MONTHLY	MONTHLY	MONTHLY	MONTHLY	MONTHLY	MONTHLY	MONTHLY	MONTHLY	MONTHLY
15	2010-01-31	784,000	-1.88			22,750	-1.73			200,000	-12.47	136,000	-8.42
16	2010-02-28	744,000	-5.10			21,000	-7.69			215,000	7.50	128,500	-5.51
17	2010-03-31	818,000	9.95			26,700	27.14			240,500	11.86	142,000	10.51
18	2010-04-30	849,000	3.79			28,400	6.37			283,000	17.67	149,000	4.93
19	2010-05-31	776,000	-8.60			25,150	-11.44			273,000	-3.53	163,500	9.73
20	2010-06-30	774,000	-0.26			25,050	-0.40			309,500	13.37	173,000	5.81
21	2010-07-31	810,000	4.65			22,500	-10.18			329,000	6.30	170,000	-1.73
22	2010-08-31	756,000	-6.67			21,100	-6.22			345,000	4.86	168,000	-1.18
23	2010-09-30	777,000	2.78			22,150	4.98			333,500	-3.33	156,000	-7.14
24	2010-10-31	745,000	-4.12			23,150	4.51			347,000	4.05	154,500	-0.96
25	2010-11-30	826,000	10.87			23,500	1.51			388,000	11.82	165,500	7.12
26	2010-12-31	949,000	14.89			24,000	2.13			391,000	0.77	168,000	1.51
27	2011-01-31	981,000	3.37			29,650	23.54			420,000	7.42	156,000	-7.14
28	2011-02-28	923,000	-5.91			28,350	-4.38			372,000	-11.43	166,500	6.73
29	2011-03-31	932,000	0.98			31,300	10.41			460,000	23.66	168,000	0.90
	Sheet1												

◆ Stock Data Management

- We assume that we already downloaded all KSE (Korean Stock Exchange) stocks' return (and price) data, KSE-monthly-return-price.xlsx.
- `my_data <- read_excel("KSE-monthly-return-price.xlsx")`
individual stock data
- `stock_name <- my_data[9,]`
- `datadate <- my_data[14:nrow(my_data),1]`

◆ DataGuide → R

	...1	Last Refresh: 2022-09-21 16:27:56	...3	...4	...5	...6	...7	...8
1	Calendar Basis	NA	NA	NA	NA	NA	NA	NA
2	Portfolio	NA	Default	NA	NA	NA	NA	NA
3	Item	NA	NA	NA	NA	NA	NA	NA
4	Frequency	월간	KRW	NA	NA	NA	NA	NA
5	Non-Trading Day	Previous	Asc	NA	NA	NA	NA	NA
6	Include Weekend	ALL	NA	NA	NA	NA	NA	NA
7	Term	20100101	Current(20220921)	NA	NA	NA	NA	NA
8	Symbol	A005930	A005930	A373220	A373220	A000660	A000660	A207940
9	Symbol Name	삼성전자	삼성전자	LG에너지솔루션	LG에너지솔루션	SK하이닉스	SK하이닉스	삼성바이오로
10	Kind	SSC	SSC	SSC	SSC	SSC	SSC	SSC
11	Item	S41000060F	S41000170F	S41000060F	S41000170F	S41000060F	S41000170F	S41000060F
12	Item Name	종가(원)	수익률(%)	종가(원)	수익률(%)	종가(원)	수익률(%)	종가(원)
13	Frequency	MONTHLY	MONTHLY	MONTHLY	MONTHLY	MONTHLY	MONTHLY	MONTHLY
14	40209	784000	-1.88	NA	NA	22750	-1.73	NA
15	40237	744000	-5.0999999999999996	NA	NA	21000	-7.69	NA
16	40268	818000	9.9499999999999993	NA	NA	26700	27.14	NA
17	40298	849000	3.79	NA	NA	28400	6.37	NA

◆ Step 1

- We first deal with dates by converting date column into Data format in R.
- Dates in R data does not look the same as the date in Excel
→ we have to convert it to dates format in R

	...1
1	40209
2	40237
3	40268
4	40298
5	40329
6	40359
7	40390
8	40421

◆ Dates

● Usual code

```
datadate1 <-  
as.Date(as.numeric(as.character(datadate)),origin="1  
899-12-30") → does not work!! → as.numeric doesn't work since  
datadate is listed.  
datadate1 <- unlist(datadate)  
datadate2<-  
as.Date(as.numeric(as.character(datadate1)),origin="1  
899-12-30")  
datadate4 <- data.frame(datadate2)
```

◆ Converted dates looks this way...

	date2
1	2010-01-31
2	2010-02-28
3	2010-03-31
4	2010-04-30
5	2010-05-31
6	2010-06-30
7	2010-07-31
8	2010-08-31
9	2010-09-30
10	2010-10-31
11	2010-11-30

◆ Sorting required data

```
ret_entire <- rbind(stock_name,  
my_data[14:nrow(my_data),]): stock data with stock names  
ret_entire <- ret_entire[,seq(3,ncol(my_data),2)]: sort stock  
returns only  
date <- rbind(c(NA),datadate4): match dimension with stock data  
ret_all <- cbind(date, ret_entire): what we need
```

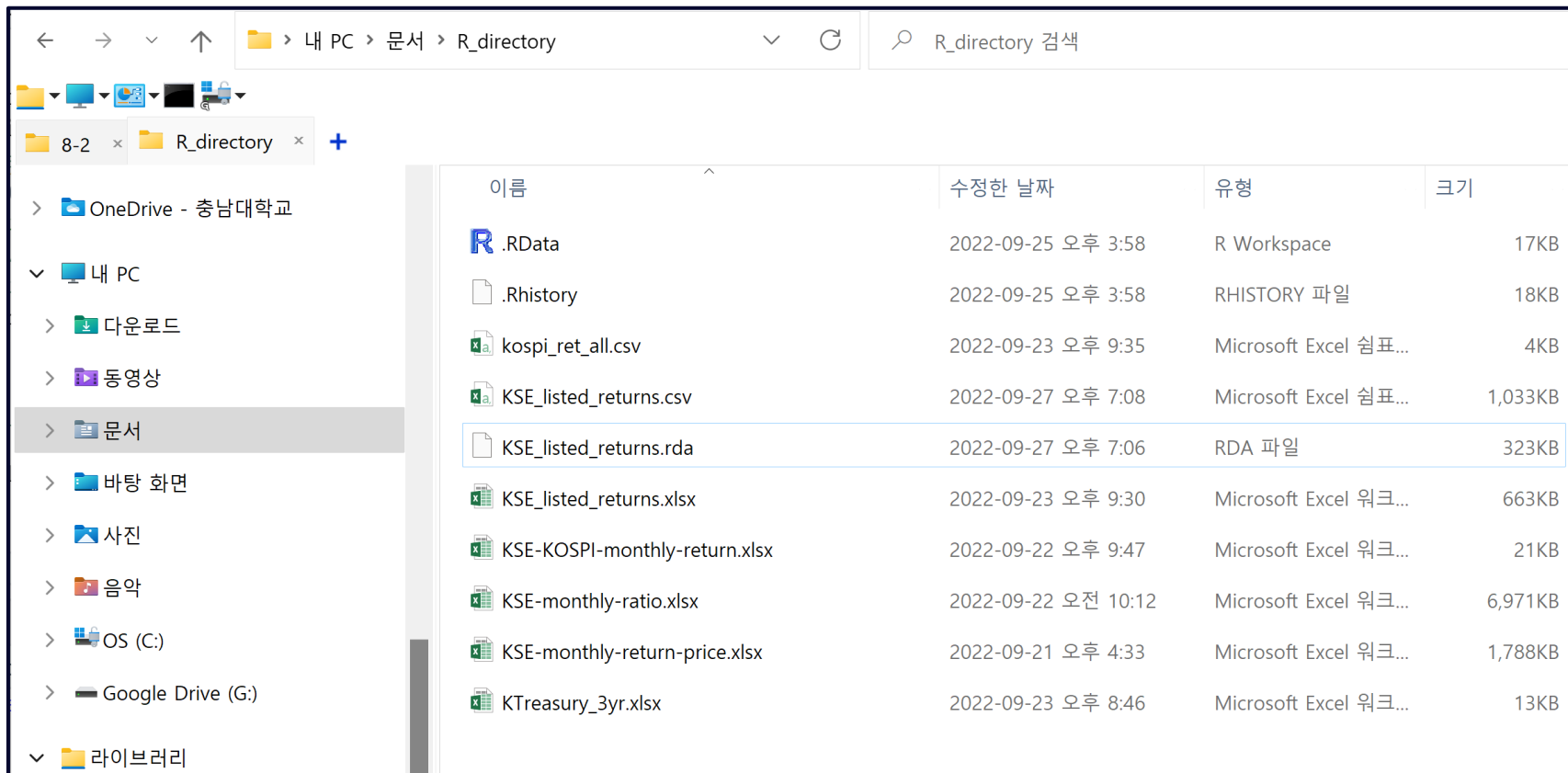

Sorted stock returns











	datadate2	...3	...5	...7	...9	...11	...13
1	NA	삼성전자	LG에너지솔루션	SK하이닉스	삼성바이오로직스	LG화학	삼성SDI
2	2010-01-31	-1.88	NA	-1.73	NA	-12.47	-8.42
3	2010-02-28	-5.099999999999996	NA	-7.69	NA	7.5	-5.51
4	2010-03-31	9.949999999999993	NA	27.14	NA	11.86	10.51
5	2010-04-30	3.79	NA	6.37	NA	17.670000000000002	4.93
6	2010-05-31	-8.6	NA	-11.44	NA	-3.53	9.73
7	2010-06-30	-0.26	NA	-0.4	NA	13.37	5.81
8	2010-07-31	4.6500000000000004	NA	-10.18	NA	6.3	-1.73
9	2010-08-31	-6.67	NA	-6.22	NA	4.8600000000000003	-1.18
10	2010-09-30	2.78	NA	4.9800000000000004	NA	-3.33	-7.14
11	2010-10-31	-4.12	NA	4.51	NA	4.05	-0.96
12	2010-11-30	10.87	NA	1.51	NA	11.82	7.12
13	2010-12-31	14.89	NA	2.13	NA	0.77	1.51

◆ Saving the data

```
install.packages("writexl")  
library(writexl)  
write_xlsx(ret_all, path="C:\\Users\\Hogyu  
Jhang\\Documents\\R_directory\\KSE_listed_returns.xlsx")  
write_xlsx(ret_all,  
path="/Users/hogyujhang/KSE_listed_returns.xlsx") # for Mac  
write_csv(ret_all, file="\\Users\\Hogyu  
Jhang\\Documents\\R_directory\\KSE_listed_returns.csv")  
save(ret_all, file="\\Users\\Hogyu  
Jhang\\Documents\\R_directory\\KSE_listed_returns.rda")
```

◆ Saved files look like...



이름	수정한 날짜	유형	크기
 .RData	2022-09-25 오후 3:58	R Workspace	17KB
 .Rhistory	2022-09-25 오후 3:58	RHISTORY 파일	18KB
 kospi_ret_all.csv	2022-09-23 오후 9:35	Microsoft Excel 실효...	4KB
 KSE_listed_returns.csv	2022-09-27 오후 7:08	Microsoft Excel 실효...	1,033KB
 KSE_listed_returns.rda	2022-09-27 오후 7:06	RDA 파일	323KB
 KSE_listed_returns.xlsx	2022-09-23 오후 9:30	Microsoft Excel 워크...	663KB
 KSE-KOSPI-monthly-return.xlsx	2022-09-22 오후 9:47	Microsoft Excel 워크...	21KB
 KSE-monthly-ratio.xlsx	2022-09-22 오전 10:12	Microsoft Excel 워크...	6,971KB
 KSE-monthly-return-price.xlsx	2022-09-21 오후 4:33	Microsoft Excel 워크...	1,788KB
 KTreasury_3yr.xlsx	2022-09-23 오후 8:46	Microsoft Excel 워크...	13KB

◆ Pick some random stocks from the stock universe

- `set.seed(100)` # For the reproducibility
- `pick_stocks <- floor(runif(10,1,ncol(ret_all)-1))` # picker index
- `ret_ind_raw <- ret_all[,pick_stocks]` # picked returns
- **## Now we remove stocks with NA values**

- **## method 1**

```
ret_ind <- ret_ind_raw[ , colSums(is.na(ret_ind_raw))==0]
```

- **## method 2**

```
ret_ind_1 <- ret_ind_raw %>% select_if(~ !any(is.na(.)))
```


◆ Picked stocks look like...

	...489	...409	...879	...89	...769	...589
1	SPC삼립	미원상사	대한제분	엔씨소프트	삼익THK	고려제강
2	-5.97	2.41	-6.57	-12.71	-2.98	-10.1
3	-0.46	-0.47	0.39	1.92	-0.82	-1.71
4	5.68	-1.06	1.95	7.89	17.98	18.670000000000002
5	1.97	0.35	12.21	16.38	94.4	-11.83
6	-5.81	5.36	-3.06	16.170000000000002	9.73	1.94
7	0.91	7.56	13.68	4.9000000000000004	1.31	-1.9
8	8.369999999999992	-4.41	-5.86	-7.62	-16.53	2.9
9	-1.98	1.0900000000000001	-6.56	22.34	-10.68	2.15

◆ Import KOSPI index returns

- We import KOSPI index returns from KES-KOSPI-monthly-return.xlsx that is downloaded from DataGuide.
- Sorting returns to suit for our analysis is exactly the same as before.

	Dates	Returns
1	2010-01-31	-4.7699999999999996
2	2010-02-28	-0.49
3	2010-03-31	6.16
4	2010-04-30	2.88
5	2010-05-31	-5.76
6	2010-06-30	3.48
7	2010-07-31	3.59