충남대학교
Chungnam National University

투자론

- *R*과 Excel을 통한 금융데이터 분석 -

8주차
R의 Collecting Data 및 Portfolio Analysis

충남대학교
장호규 교수

충남대학교
Chungnam National University

Unit 01

Collecting Data with R

충남대학교
Chungnam National University

Overview

- Using API for collecting data
- Crawling

API

- API stands for Application Programming Interface. In the context of APIs, the word Application refers to any software with a distinct function. Interface can be thought of as a contract of service between two applications. This contract defines how the two communicate with each other using requests and responses.

- API can be widely used in many tasks… such as collecting data

- There are two ways to collect data with an API in R (and Python).
  1) Use a library that comes packaged with functions that call the API. This is by far the easiest. 2) If you can't find a library that makes calls to the API of interest, then you need to make direct calls to the API yourself.

Case Study

Google Trends

```
install.packages("gtrendsR")
library(gtrendsR)
library(tidyverse)
```

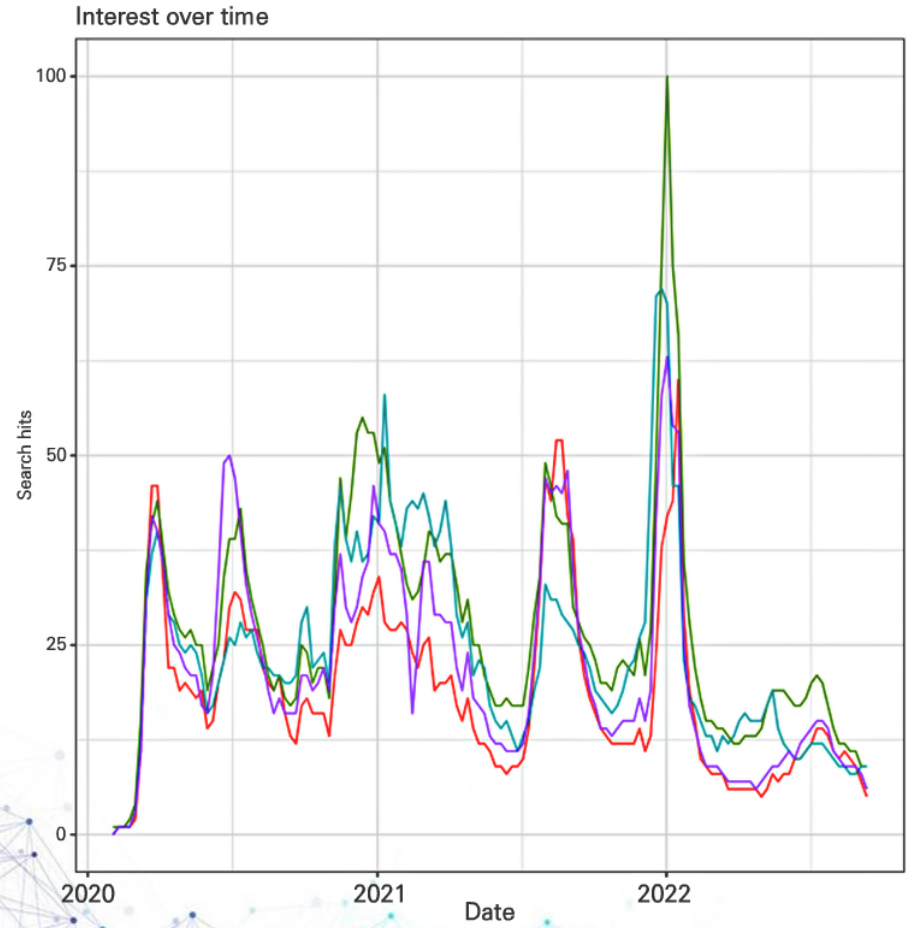
- `## get web query activity for keyword = "covid" for queries`
- `## originating in states of California, Texas, New York and Alabama`

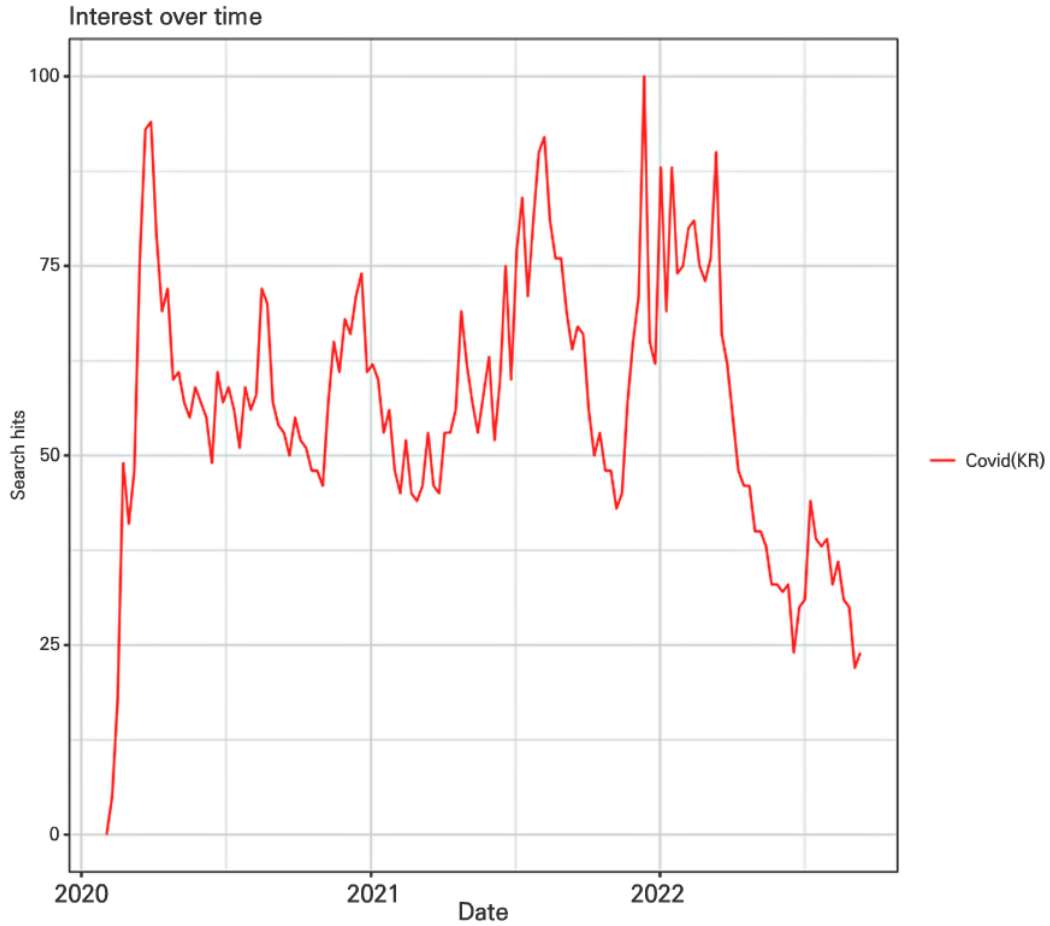
```
res <- gtrends(c("covid"), time="2020-02-01 2022-08-31",
geo=c("US-CA", "US-TX", "U-NY", "US-AL"))
```

- For the detailed information, refer to **https://cran.r-project.org/web/packages/gtrendsR/gtrendsR.pdf**

Case Study

Google Trends





getSymbols()

- Yahoo Finance offers stock price data for free.
- Quantmod package in R provides `getSymbols()` function that can access Yahoo Data API to download the relevant data.
- library(quantmod)
- `getSymbol('AAPL')` → 'xts' format (check: `class(AAPL)`)

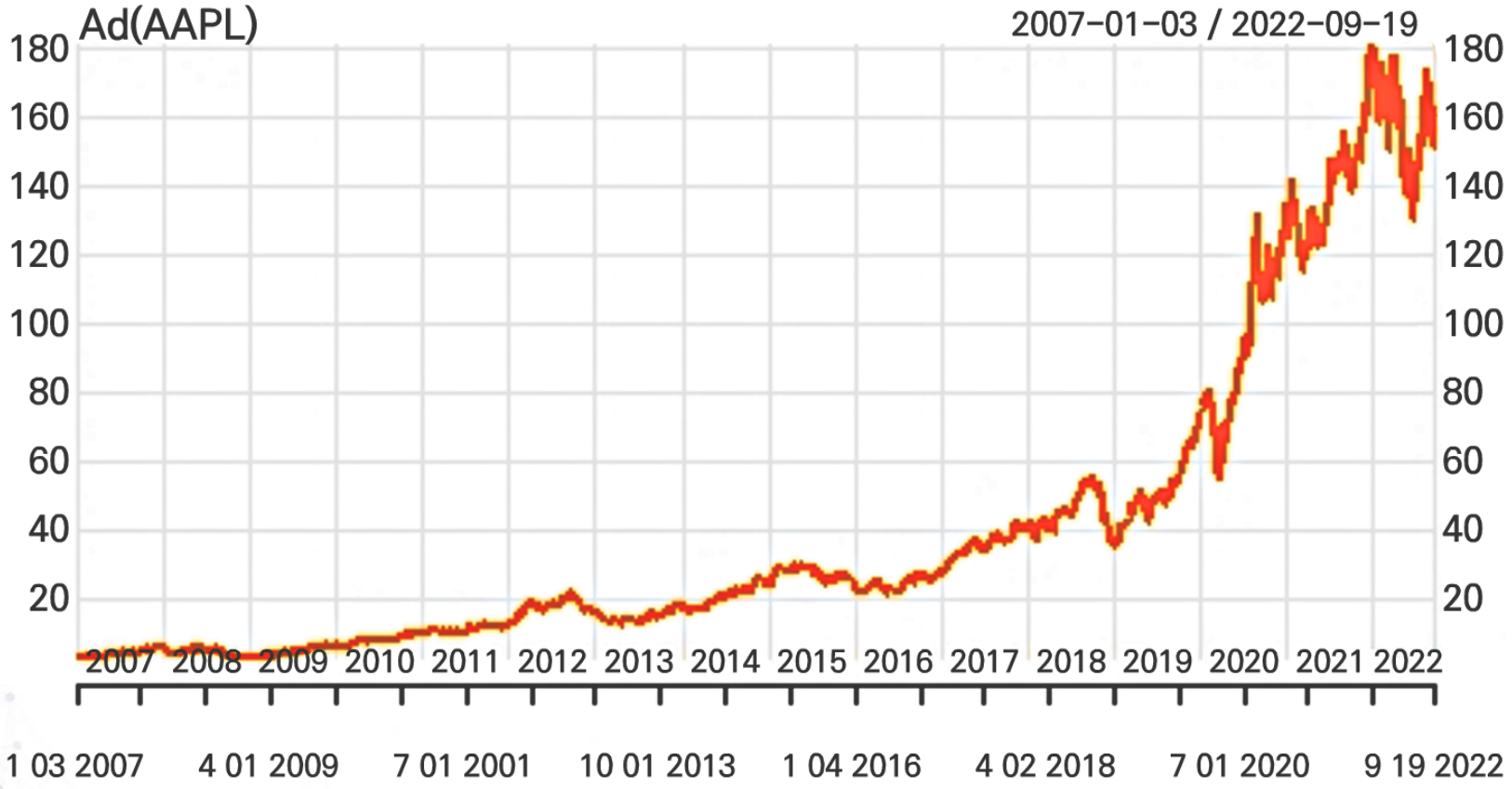
"YYYY-MM-DDTHH:MM:SS"

```
> head(AAPL)
           AAPL.Open AAPL.High AAPL.Low AAPL.Close AAPL.Volume AAPL.Adjusted
2007-01-03  3.081786  3.092143 2.925000   2.992857  1238319600      2.555398
2007-01-04  3.001786  3.069643 2.993571   3.059286   847260400      2.612116
2007-01-05  3.063214  3.078571 3.014286   3.037500   834741600      2.593514
2007-01-08  3.070000  3.090357 3.045714   3.052500   797106800      2.606321
2007-01-09  3.087500  3.320714 3.041071   3.306071  3349298400      2.822829
2007-01-10  3.383929  3.492857 3.337500   3.464286  2952880000      2.957917
```

출처 : 교수자 제공

`chart_Series(Ad(AAPL))`

- Ad() collects adjusted price



Another Method

```
data = getSymbols('AAPL',
                  from = '2000-01-01', to = '2018-12-31',
                  auto.assign = FALSE)

Head(data)
```

	AAPL.Open	AAPL.High	AAPL.Low	AAPL.Close	AAPL.Volume	AAPL.Adjusted
2000-01-03	0.936384	1.004464	0.907924	0.999442	535796800	0.853355
2000-01-04	0.966518	0.987723	0.903460	0.915179	512377600	0.781409
2000-01-05	0.926339	0.987165	0.919643	0.928571	778321600	0.792843
2000-01-06	0.947545	0.955357	0.848214	0.848214	767972800	0.724232
2000-01-07	0.861607	0.901786	0.852679	0.888393	460734400	0.758538
2000-01-10	0.910714	0.912946	0.845982	0.872768	505064000	0.745197

출처 : 교수자 제공

Multiple Download

```
ticker=c('FB', 'NVDA')
getSymbols(ticker)
```

- **Domestic Stocks – a case of Samsung**

```
getSymbols('005930.KS', from='2010-01-01', to='2022-09-20')
tail(Ad(`005930.KS`))
```

- **For domestic stocks, it is better to use closing price instead of adjusted one due to errors**

```
tail(Cl(`005930.KS`))
```

- **Celtrion Case:**

```
getSymbols("068760.KQ", from='2010-01-01' to='2022-09-20')
```

FRED (Federal Reserve Economic Data)

- **US Treasury 10-yr**

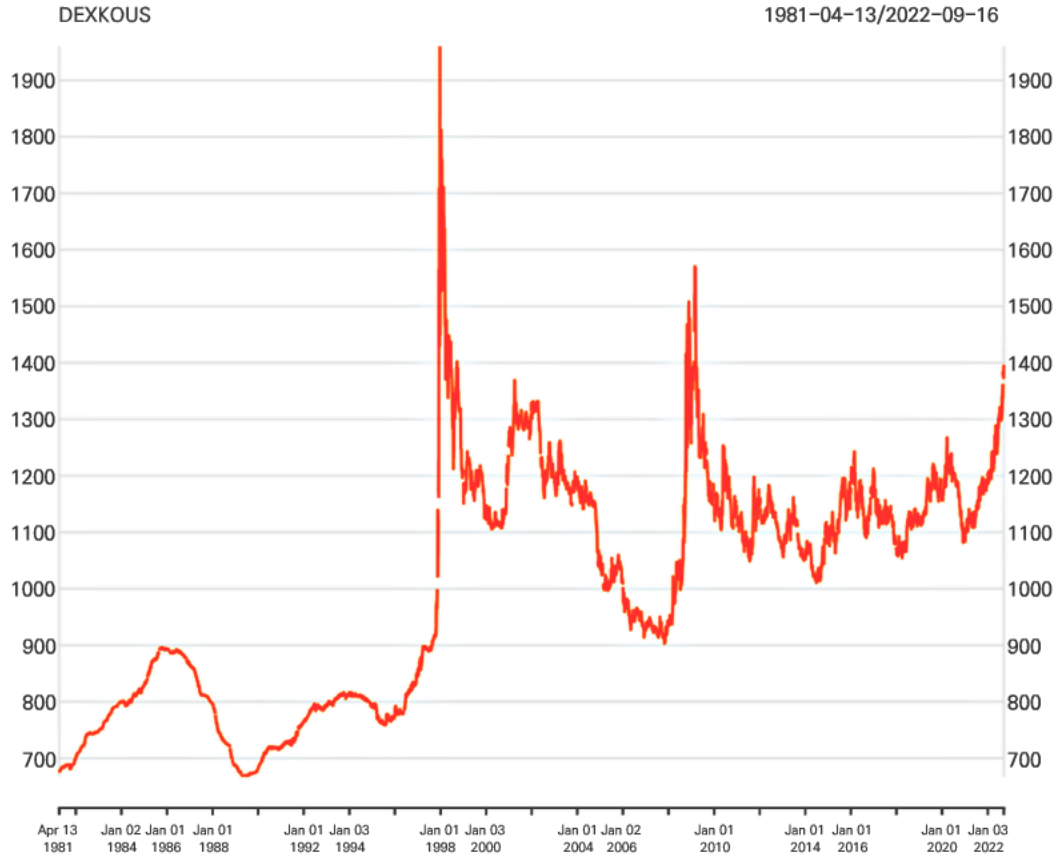
```
getSymbols('DGS10', src='FRED')
chart_Series(DGS10)
```

- **How to find a ticker: access FRED to search**

  → ex) Korean exchange rate (dollar/won): DEXKOUS

```
getSymbols('DEXKOUS', src='FRED')
chart_Series(DEXKOUS)
```

FRED (Federal Reserve Economic Data)



Method of Collecting Financial Data

- Use API to crawl the data
- Two methods
  : GET (based on internet address) & POST (based on user interface)
- We don't go further into crawling. Rather we use the data that is already collected.
- Thus, we focus on the data management and apply portfolio optimization method.