



Sungkyunkwan University (SKKU) International Summer Semester (ISS) 2021

## **Introduction to Data Science for Investment**

**Prof. Youngju Nielsen, Graduate School of Business, Sungkyunkwan University**

### **SHORT COURSE DESCRIPTION**

As world financial markets become increasingly complex, we are demanded to analyze the continuous flow of information. Quantitative models based on statistical foundations and programming skills are vital ingredients to process information accurately and promptly. In this course, we use R programming to collect and analyze the data, optimize the investment portfolio, and trade securities. In the earlier part of the course, you will learn the essential components of R programming. The course presumes no prior programming knowledge and experience. Therefore, R programming instruction starts with the installation of the software.

Then we will improve our skills in data analysis and programming by doing the most accepted practice in investment analysis. You will start working on analyzing the past returns and risk and then extend this to statistical modeling to use a factor model. Then finally, you will use optimization to construct the portfolio.

The course is well suited for a wide range of audiences. If you want to know how to program using R, you will learn that with the investment examples. If you're going to have extensive quantitative skills in the investment field, this course can be the first step. If you just want to use data analysis for your future personal investment, this course will give you enough tools to start.

Investment requires significant knowledge in financial economics, general economics, statistics, and mathematics. However, this course will not focus on studying the theories but on implementing the concepts.

Students completing this course will be able to do the following:

---

- Understand the big frame of the analytic team's tasks – acquiring data, analyzing sizable data using programming, and automating the analytic process.
  - Do computer program coding.
  - Deal with a bigger size (beyond the volume of data MS Excel can deal with) data.
  - Analyze investment's return and risk without tedious copy and paste in MS Excel.
  - Build your investment strategy.
- 

### **READING MATERIALS**

---

There is no required textbook. I will post the complementary materials (readings, websites, and videos) .

You will use your laptop during class. All class instruction will be given based on the Windows O/S version.

Minimum requirements for a laptop

- Minimum 8GB RAM
- Enough disk space for data, recovered files, image files etc.
- The administrative privilege is required to install and run utilities.
- A network connection

You will find more information from the link below.

[https://cran.r-project.org/bin/windows/base/rw-FAQ.html#Can-I-use-R-on-64\\_002d-bit-Windows\\_003f](https://cran.r-project.org/bin/windows/base/rw-FAQ.html#Can-I-use-R-on-64_002d-bit-Windows_003f)

[https://www.r-studio.com/Unformat\\_Help/systemrequirements.html](https://www.r-studio.com/Unformat_Help/systemrequirements.html)

### **COURSE REQUIREMENTS AND GRADING**

---

Participation & attendance	20 points
Team activity during the classes	20 points
Homework (Quiz in the 2 <sup>nd</sup> week is included in the Homework category.)	20 points
Written exam	20 points
Final project presentation	20 points

- Written exam (20 points)

There will be one one-hour exam. The exam questions will test whether you understand the concepts. As long as you reviewed the class material, you will not have any difficulty to answer questions.

- Final project presentation (20 points)

You will never learn anything in analytics and programming unless you make your hands dirty! You will use everything you learned from the course for this final project. This course's final project is teamwork.

The detail of the project guideline will be announced in the first week.

- Team tasks (20 points)

The professor will give the same data analysis tasks to teams during the classes. Your team will compete for prompt decision makings and thorough analyses. The scores will be given according to your team's performance each time.

- Participation & attendance (20 points)

Although this course is somewhat technical, each student is expected to take part in the in-class discussion. Active class participation will become more comfortable if you master the previous lecture's content and finish homework.

- ✓ Top grades on this component assume the class participants do the following:
  - ✓ Attending all class sessions
  - ✓ Sharing the solutions and tips found while you are doing homework
  - ✓ Showing evidence that students have been practicing programming in R.
  - ✓ Raising the fundamental questions to be worth to analyze to the instructor and fellow students
  - ✓ Contributing positively to the investment case discussion

To get A from this component, you are expected to do all the above in every class.

To get B+ from this component, you are expected to do all the above in 2/3 of the classes. It would be best if you also had answered questions appropriately when called on in class.

I would like to see B+ and above as the average grade for this component.

## COURSE SCHEDULE

---

Date	Session	Subject	Detail
Week1: Getting ready with R programming			
Thursday (June 24th)	1	Course introduction (RT)	<ul style="list-style-type: none"> <li>• Explanation</li> <li>• R and R Studio installation</li> </ul>

	2	Basic programming in R	<ul style="list-style-type: none"> <li>Dealing with different forms of data</li> </ul>
Friday (June 25th)	1	The data structure in R	<ul style="list-style-type: none"> <li>Vector, matrix, factor, list, and data frame</li> </ul>
	2	Essential grammar in R	<ul style="list-style-type: none"> <li>If, for, while</li> <li>Making teams who will work together over week 2 to week 4 (the professor will lead the team forming activity. Students do not need to decide teams.)</li> </ul>
Week2: Investigating what happened to the Disney and Apple stocks			
Monday (June 28th)	1	Analyzing past performance	<ul style="list-style-type: none"> <li>Reading csv file data sets to R and do a descriptive analysis</li> </ul>
	2	Team activity Quiz covering the week1 material	<ul style="list-style-type: none"> <li>Working on team's stock choices</li> </ul>
Tuesday (June 29th)	1	Analyzing factors	<ul style="list-style-type: none"> <li>Reading the data directly from websites and make a factor model using a linear regression model</li> </ul>
	2	CAPM theory and modeling theory	<ul style="list-style-type: none"> <li>Understanding economic and financial meaning of factors deeply</li> </ul>
Wednesday (June 30th)	1	Forecasting future returns	<ul style="list-style-type: none"> <li>Setting up different analysis and modeling structure for predictive models (vs. descriptive models)</li> </ul>
	2	Team activity	<ul style="list-style-type: none"> <li>Working on a forecasting model</li> </ul>
Thursday (July 1st)	1	Comparing investment strategies	<ul style="list-style-type: none"> <li>Making investment strategies</li> </ul>
	2	Team activity & presentation	<ul style="list-style-type: none"> <li>Presenting and suggesting what the team will invest based on past performance and future performance forecasts</li> <li>Homework1(Due July 6th): Get S&amp;P500 data, divide the</li> </ul>

			dataset into before 1990 and after 1990. Find the relationship with dividend yields using historical data.
Week3: Understanding the characteristics of stocks using returns and risks			
Tuesday (July 6th)	1	Analyzing the past performance using factors	<ul style="list-style-type: none"> <li>Figuring out how factors affected performance</li> </ul>
	2	Homework 1 Presentation & Discussion	<ul style="list-style-type: none"> <li>Each team will present about 10-15 minutes</li> </ul>
Wednesday (July 7th)	1	Comparing stocks based on risks	<ul style="list-style-type: none"> <li>Learning two different ways of calculating risks</li> </ul>
	2	Team activity	<ul style="list-style-type: none"> <li>Discussing two different stocks' risk profile.</li> </ul>
Thursday (July 8th)	1	CAPM and factor models	<ul style="list-style-type: none"> <li>Modeling risks (instead of calculating from the past data)</li> </ul>
	2	Team activity	<ul style="list-style-type: none"> <li>Making the risk model for two stocks.</li> </ul>
Friday (July 9th)	1	Introducing tidyvers package	<ul style="list-style-type: none"> <li>Learning one of the most popular R packages, tidyvers and ggplot2.</li> </ul>
	2	Team activity	<ul style="list-style-type: none"> <li>Drawing the analysis conclusion from graphs</li> <li>Homework2 (Due July 12<sup>th</sup>) : Compare DIS and AAPL using factor models and make meaningful graphs. Write what you learned from your analysis.</li> </ul>
Week4: Constructing investment portfolios based on returns and risks			
Monday (July 12 <sup>th</sup> )	1	Finding the data to work with®	<ul style="list-style-type: none"> <li>Sourcing free online data from various websites and databases</li> </ul>
	2	Homework 2 Presentation & Discussion	<ul style="list-style-type: none"> <li>Each team will present about 10-15 minutes</li> </ul>
Tuesday (July 13th)	1	Portfolio optimization	<ul style="list-style-type: none"> <li>Making optimization code</li> </ul>
	2	Portfolio optimization theory	<ul style="list-style-type: none"> <li>Understanding the economic, financial, and mathematical meaning of mean-variance optimization</li> </ul>

Wednesday (July 14th)	1	Making graphs and summary of portfolios	<ul style="list-style-type: none"> <li>Analyzing the optimized portfolio in detail and presenting it to investors</li> </ul>
	2	Team activity	<ul style="list-style-type: none"> <li>Making your team's optimized portfolio</li> </ul>
Thursday (July 15th)	1	Final written exam	
	2		Final project report (Due July 20th):