This course will be offered in English. The goal of this course is to help students to handle financial data using R computer language. More specifically, the course will help students to write R programs to conduct financial analytics, while the purpose of financial analytics is to transform financial data to insights (through visualization in particular) that support, improve, and automate financial decisions. The majority of class time will be used to explain various R language skills so that students can eventually use R to do a real-world financial analysis.

All the teaching materials will be provided by the teacher through a computer software called PresTree, which runs only on Windows PC. It will be a problem for those students who have Apple or Linux computers. These students should either try to install a Windows subsystem in their computers to run PresTree or to buy/borrow a Windows PC. Also, it is usually helpful to have a computer with a touch-screen so that class notes can be directly written on the PresTree screen.

"The big data" in the course title only refers to time-series and cross-sectional datasets of large sizes. Unstructural data such texts, images, etc., will not be dealt with. This course will not cover artificial intellectual subjects either and only expose to very limited machine learning techniques.

The course grade will be based on a number of in-class quizs/homeworks (40%) and a final report (60%). The subject of the final report however is not limited to (but is encouraged to) finance areas but can be various social issues. A proposal for your final report should be turned in three weeks before the end of this semester. You will then be advised whether your final report is feasible and acceptable.

You should install R and a companion software R-Studio in your computers before class begins. The installment process is simple and straightforward. You should search for solution on the internet if any problem occurs.

The main topics to be covered in the course:

- R language basics

  1. Operation units: vectors, lists, factors, matrices, data-frames,
  2. Presentation skills: R markdown, shiny
  3. Control structures: conditional statements, iteration statements, writing functions
  4. Using packages
  5. Getting datasets
  6. Statistical operations
     - Statistical functions
     - Handling missing values
     - Simulation
     - Matrix algebra
  7. Basic plots: scatter plots, line plots, bar plots and tables, histogram, box plots

- Tidyverse ecosystem

  1. Special data-frames: tibbles
  2. Data wrangling: dyplr packages
     - Selecting variables, filtering observations, defining variables, merging data
     - subsample data wrangling, linear regressions, and pivot-tables
  3. Data-visualization: ggplot2 packages
     - The main structure: aesthetics and facetting
     - Layers: scatter, line, functional, bar, histogram, and box plots; summary statistic plots
     - Themes
     - Axes and coordinate systems: scale functions
     - Legends: guide functions
  4. Iterative function execution: purrr packages
  5. Pivot Table technique: tidyr package
     - Wide-forms and long-forms
  6. String manipulation: stringr package
     - Query, substring, parsing, ordering, concatenation, plugging, space handling, splitting, searching, replacing
     - Regular expressions
  7. Factor manipulation: forcats package

- Time-series analysis:

  1. Date classes: formats, series, and lubridate package
  2. Time-series classes:
     - ts series

- zoo/xts series
- tsibble series

3. Forecasting: fable package

- ARIMA models
- VAR Models
- GARCH models
- ETS models
- Linear regression models

4. Series smoothing: MFilter package

5. Detecting anomalies: anomalize package

6. Hypothesis testing: ACF, Unit-root testing, normality testing,

- Financial data analysis

1. Downloading and plotting financial data: quantmod package

2. Measuring the performance of financial returns: PerformanceAnalytics and tidyquant packages

- Returns calculations and backtesting
- Sharpe ratios
- CAPM
- Drawdowns
- Downside risks
- VaR, ES

3. Portfolio optimization: PortfolioAnalytics package

- Constrained portfolio optimization
- Constrained portfolio optimization with rebalancing
- Constrained random portfolios

If you know something about R, then you will realize there are a lot of materials to be covered in the class. Because of this, all the topics will be taught only briefly in the class but will be thoroughly tested either in quizs or homeworks so long as they are included in the PresTree software. This class will be quite time-consuming if you want a good grade.