# Course Introduction and Syllabus

# Zhejiang University

**Course code：** 2023112 **Course title：** Programming for Data Analysis

**Credit：** 2 **Teaching hour：** 32

**Professor：** CAO Zike

**Prerequisites：** Basic knowledge about computer programming

## Course Introduction

As a free and powerful open source language, R is developed for data exploration, statistical computing and graphics application. In this era for data science, R has become an efficient, elegant and widely adopted tool for statistics, econometric analysis and data mining both inside corporations and academia. Giant companies such as Google, Pfizer, Merck, Bank of America and Shell are using it to cope with their massive data. After learning this course, the students will acquire the knowledge of how to effectively use the power of R to tackle with data analysis problems with datasets from the real world and have a better understanding of the principle of statistical computing and econometric analysis. In this course, we will systematically introduce the basic knowledge of R, data import, data arrangement and manipulation, statistical simulation, function writing, statistical description a linear regression, as well as the data visualization with the powerful function of R for graphics.

## Learning Objectives

1) Be able to import/export, manipulate data and run effective statistical computing with R.
2) Acquire the skills to do the programming and simulation in R.
3) Know how to illustrate the data with graphics by virtue of the powerful visualization function of R.
4) Know how to conduct factor analysis in R.

## Measurable Learning Outcomes

1) Do data preprocessing according to research requirements.

2) Use plots to present basic and advance graphs.
3) Be able to report the basic descriptive statistical results.

## Teaching Schedule

Week 1: Introduction to R: Brief introduction about R and basics of data structures
Week 2: Data visualization using ggplot2
Week 3: Data transformation using dplyr
Week 4: Exploratory data analysis using ggplot2 and dplyr
Week 5: Relational data using dplyr Data reshaping using tidyr
Week 6: Writing Your Own Functions Iteration Loop and Functional programming
Week 7: Probability and Regression
Week 8: In class Examination