

# R PROGRAMMING AND APPLICATIONS TO FINANCE

## Lecture 1: Introduction and Data Manipulation

Master of Arts in Finance

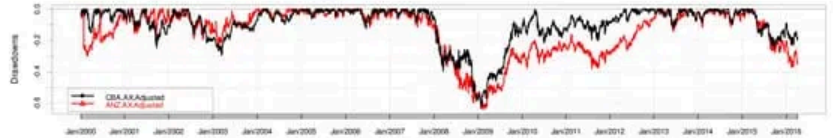
International College of Economics and Finance

The Finance and Financial Planning HDR Research Methodology Workshops

Time: 10:00-11:00am, 13th May 2016  
Place: NSO 0.32 Griffith University

**R** in Finance  
Liang C. Zhang

Introduction to R and Its Applications in Finance  
Slides of this presentation: <https://goo.gl/4oPlja>  
R script for this presentation: <https://goo.gl/2YKQpq>



# COURSE OVERVIEW

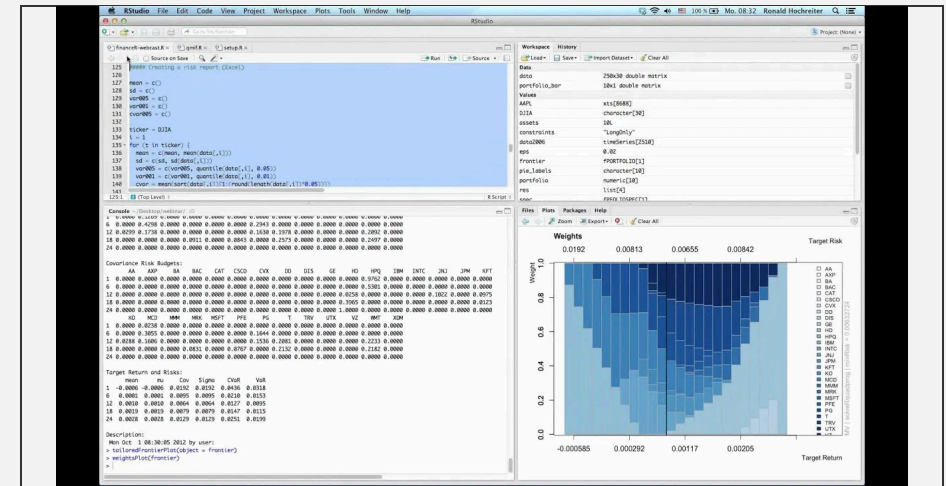
Welcome to **R Programming and Applications to Finance**. This course bridges financial theory and practical implementation for MA students.

## Core Goals

- Master R for financial data analysis
- Implement advanced volatility and risk models
- Optimize portfolios using real-world data

## Prerequisites

Solid background in statistics and econometrics. Basic understanding of financial modeling concepts.



## Focus Areas

- Data Manipulation (Tidyverse)
- Time Series Analysis (ARIMA/GARCH)
- Market & Credit Risk Management

# R VS. PYTHON IN FINANCE

FEATURE	R (THE SPECIALIST)	PYTHON (THE GENERALIST)
Statistical Depth	Unparalleled for niche financial models (GARCH, Bayesian).	Strong, but often lags in specialized econometrics.
Visualization	`ggplot2` offers a superior "Grammar of Graphics".	`matplotlib` and `seaborn` are powerful but less intuitive.
Data Handling	`data.table` and `tibble` are built for data frames.	`pandas` is the standard, inspired by R's data frames.
Community	Academic, research, and high-end quant finance.	Software engineering, AI, and general data science.

*"R is built by statisticians for statisticians; Python is a general-purpose language that does statistics."*

---

# FLAVORS OF R

## BASE R

The core language and its built-in functions.

- **Stability:** Code written 20 years ago still runs today.
- **No Dependencies:** Lightweight and always available.
- **Syntax:** Can be inconsistent and verbose for complex tasks.

## TIDYVERSE

A collection of packages for modern data science.

- **Readability:** Uses the pipe operator for intuitive workflows.
- **Consistency:** Shared design philosophy across all packages.
- **Ecosystem:** Seamless integration with plotting and modeling.

## DATA.TABLE

High-performance version of data frames.

- **Speed:** Extremely fast for large-scale financial data.
- **Memory:** Highly efficient memory management.
- **Syntax:** Concise but has a steeper learning curve.

---

# WHY TIDYVERSE?

---

## HUMAN-READABLE CONSISTENCY

Uses the pipe operator `%>%` or `|>` to chain operations like a sentence. Code becomes a logical sequence rather than nested functions.

Functions share a common design and data structure (tibbles). Once you learn one package, others feel intuitive.

---

## SEAMLESS INTEGRATION

Works with `ggplot2` for visualization and `broom` for tidying model outputs to create a unified workflow.

---

## HIGH PRODUCTIVITY

Faster to write, easier to debug, and ideal for reproducible research. Analysts focus on insights rather than syntax.

**"The Tidyverse is the modern standard for data science in R."**

---

# DATA MANIPULATION WITH DPLYR

---

## filter()

Pick observations by their values. Essential for selecting specific time periods or assets.

```
data %>% filter(Return > 0)
```

---

## summarize()

Collapse many values into a single summary. Compute mean, SD, or VaR.

```
data %>% summarize(Mean = mean(Ret))
```

---

## select()

Pick variables by their names. Isolate prices, volume, or specific indicators.

```
data %>% select(Date, Price)
```

---

## arrange()

Reorder the rows. Sort by date or by performance metrics.

```
data %>% arrange(desc(Date))
```

---

## mutate()

Create new variables. Calculate returns, moving averages, or risk metrics.

```
data %>% mutate(LogRet =  
  log(P/lag(P)))
```

---

## The Pipe ( % > % )

Chain multiple operations together for readable, sequential data processing.

```
data %>% filter(...) %>% mutate(...)
```

# LONG VS. WIDE DATA FORMATS

## WIDE FORMAT

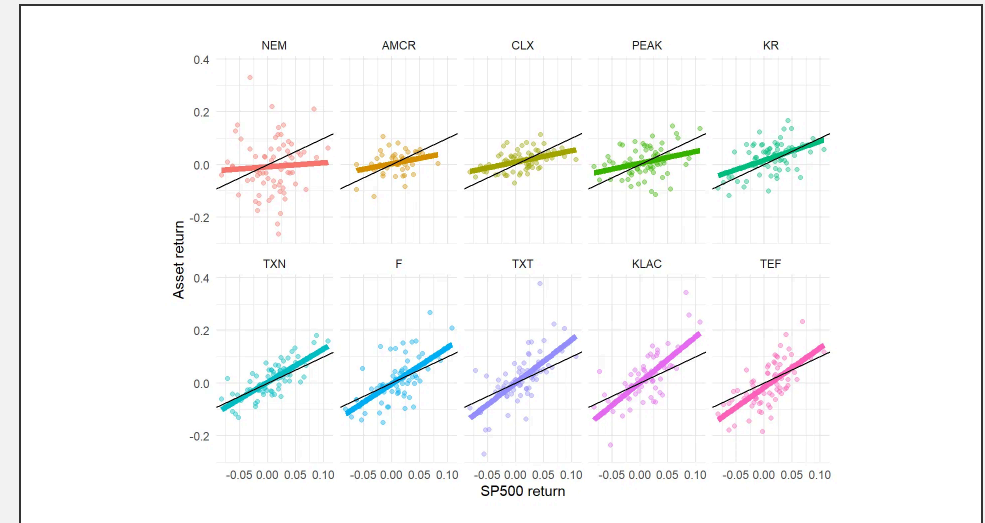
Each variable is a column. Common in spreadsheets and traditional financial reports (e.g., columns for AAPL, MSFT, GOOG prices).

## LONG FORMAT (TIDY)

Each observation is a row. Preferred by the Tidyverse for plotting with ggplot2 and statistical modeling.

## TRANSFORMATION TOOLS

- `pivot_longer()`: Wide → Long
- `pivot_wider()`: Long → Wide



*Financial time series data often requires transformation to long format for multi-asset plotting.*

## WHY IT MATTERS

Tidy data principles ensure that each variable forms a column and each observation forms a row, simplifying the analysis pipeline.

---

# FINANCIAL TIME SERIES DATA

---

## ts

The **Base R** time series object. Best for regular intervals (monthly, quarterly). Limited support for irregular dates or multiple time zones.

## xts

"Extensible Time Series". Built on top of zoo, it is the **industry standard** for financial data in R. Highly efficient for high-frequency data.

## zoo

"Z's Ordered Observations". Designed to handle **irregular time series**. Provides a flexible foundation for many other packages.

## tsibble

The **modern tidyverse** approach. Integrates time series analysis with tidy data principles, making it ideal for forecasting and visualization.

*"Choosing the right data structure is critical for efficient financial modeling and backtesting."*



---

# PLOTTING WITH GGPLOT2

---

## GRAMMAR OF GRAPHICS

A systematic way to build plots by mapping data to aesthetic attributes (x, y, color) and geometric objects (lines, points).

---

## THE LAYERING SYSTEM

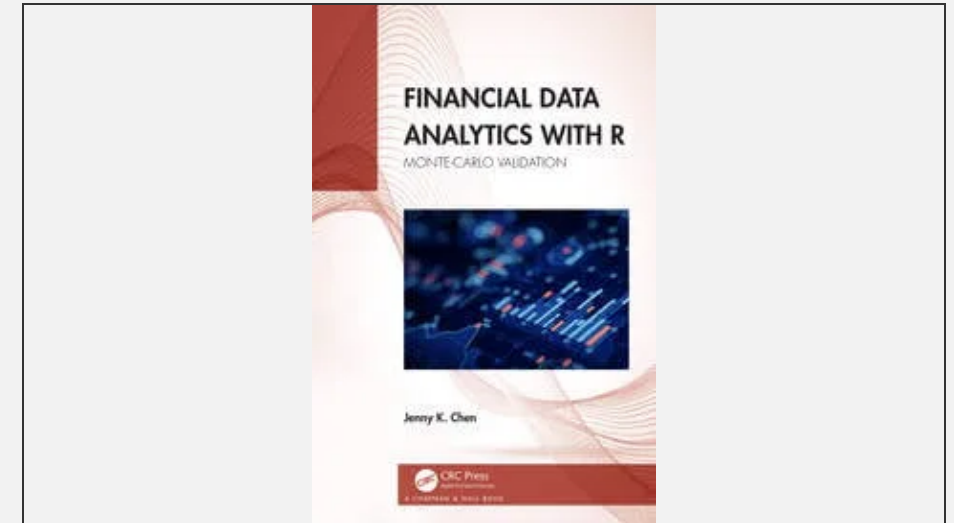
Start with data, then add layers using the + operator. This modular approach allows for complex, multi-asset visualizations.

```
ggplot(data, aes(x=Date, y=Price)) + geom_line()
```

---

## CUSTOMIZATION

Use `labs()` for titles and `theme_minimal()` for professional, publication-quality financial charts.



*Visualizing complex financial relationships requires a robust and flexible plotting system.*

---

## FINANCIAL EXTENSIONS

- tidyquant: Candlestick & OHLC charts
- ggthemes: Bloomberg & WSJ styles
- highcharter: Interactive widgets

---

# SUMMARY & NEXT STEPS

## LECTURE RECAP

---

### R in Finance

A specialized tool for statistical depth and superior financial data visualization.

### The Tidyverse

The modern standard for readable, consistent, and productive data science workflows.

### Data Foundations

Understanding Long vs. Wide formats and specialized time series objects (xts, zoo).

## UPCOMING TASKS

---

- **Hands-on Seminar:** Practice data cleaning, manipulation with dplyr, and basic plotting.
- **R Notebook:** Review the provided code examples for real-world financial data.
- **Home Assignment:** Analyze a stock's performance using the tools learned today.
- **Preparation:** Ensure R and RStudio are installed with the tidyverse package.

---

## Questions?

See you in the seminar!